SIMULATION-BASED INFERENCE AND DATA COMPRESSION APPLIED TO COSMOLOGICAL PROBLEMS

FRANCESCA GERARDI

A dissertation submitted in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

University College London

Department of Physics and Astronomy

Supervisors:

Prof. Benjamin Joachimi

Dr. Andreu Font-Ribera

Dr. Stephen Martin Feeney

Examiners:

Prof. Florian Beutler

Prof. Jason McEwen

January 22, 2024

I, Francesca Gerardi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

At this time, for cosmology, it is crucial to develop robust inference frameworks that will improve our understanding of the standard cosmological model, shedding light to open problems it currently suffers of. However, the increasing complexity of current analyses urged the need for optimal data compression algorithms and alternative simulation-based Bayesian approaches, which development was boosted by Machine Learning advances. The work presented in this thesis focuses on testing and applying these techniques to problems in the field of gravitational waves and the Lyman-α forest, and the content can be effectively split up into two main building blocks.

The first part aims at addressing whether density estimation simulation-based inference yields unbiased estimates of cosmological parameters in the presence of selection effects. As a test case we use mock binary neutron stars mergers catalogues for the Hubble constant estimation, given a toy hierarchical model. Not only did this method yield statistically unbiased estimates of H_0, but its precision almost matched the one of standard Bayesian analysis.

The second part of my work explores if and how information can be optimally and efficiently extracted from Lyman-α correlation functions. First, we aim to understand whether the baryon acoustic oscillations peak alone, as considered in standard analyses, constitutes a sufficient summary to capture all the relevant cosmological information. Performing a direct fit to the full shape of simple mock correlations, we demonstrated that there is extra information and we traced it back to the Alcock-Paczyński effect and redshift space distortions. Finally, in another work, we apply score compression to realistic mocks, finding good agreement with the traditional approach at the posterior level. Moreover, we find that the covariance matrix estimated from data via subsampling is a good approximation to the true covariance.

Impact statement

This thesis work exploits extensions to standard analysis frameworks, to increase the efficiency in cosmological inference tasks. The novel work here presented is made of two main parts, which resulted in incredibly different contributions.

The first part refers to applying neural density estimation in the field of gravitational waves. Being based on a simple hierarchical model of light standard sirens catalogue, this work does not aspire to bring a direct contribution at the real data level, but rather at the methodological level. This work sets itself as a proof-of-concept to demonstrate that in presence of selection effects, the results of simulation-based inference tools are not biased with respect to standard approaches. Demonstrating this is extremely important because of the problems traditional Bayesian analyses have to deal with, which makes simulation-based methods so appealing. Not only the parameters-data space is increasing, but accounting for selection effects in the analytical form of a likelihood can require computationally expensive and/or inaccurate approximations. Hence, proving these novel methods are reliable is a must.

The second part, which relates to Lyman-α correlation functions, instead brought contributions to the VEGA code, used within the Dark Energy Spectroscopic Instrument (DESI) Lyman-α working group for fitting and modelling the correlation functions, in terms of extensions to the standard analysis. In particular, in the directions of both performing direct cosmological inference, building an interface to the COBAYA sampler, and compressing the data vector into an optimal set of summary statistics, to increase the efficiency of the analysis.

This work led to two first author peer-reviewed publications. Beyond academia, this work faces challenges of modern cosmology which are more broadly extended. The rapid advancement in technology and the definition of Big Data structures is driving multiple fields to explore and apply these same cutting-edge tools for different purposes.

Acknowledgements

To begin with, all my gratitude goes to my supervisors Benjamin Joachimi, Andreu Font-Ribera and Stephen M. Feeney, for always being supportive along this journey, either inside or outside of academia. You gave me the chance to understand and explore what I really liked, which is priceless and I could never be more thankful for. To you also goes all my admiration, as you really showed me what being a good scientist really means, either from the academic or interpersonal perspectives. You shared with me valuable knowledge and I have always felt free to express my thoughts and say something wrong, for you to patiently guide me through problems and clear out my doubts.

Thanks to Andrei Cuceu, for always being present when asked for help and feedback, you are definitively the best collaborator anyone could ever ask for. To Justin Alsing, Seshadri Nadathur and Pablo Lemos for your precious suggestions and contributions to my research projects.

To the team I have been working with during my internship at Google. To Angus, for making such an experience the best it could have ever been. You've been a mentor outside of academia and you have shown great dedication to your work (and to coding). You have been a great teacher!

Thanks to all the people at UCL who made this experience unforgettable. Thanks to Davide and Fede, for the late McDonald's sessions and for being amazing psychologists throughout my PhD, from any location on Earth. To Cata, for being one of the most genuine and determined person I have ever met. To Pascal, Nik and Maria, for always being there when in need for beer(s) or a chat. To Xie, you've been my very first friend at UCL and my source of fries and fried chicken during these years, they were fuel to this work. To Kiyam, for sharing with me the bitter sweet joys of Tensorflow. To Lilian, for bearing my question about your food recipes. To Max and Claudia for being great office mates and even better at playing Mario Kart. To Alessio, Chris and Alex, for the great advises you gave me, from your postdoc wisdom.

To Andrea, for supporting me in the decision of going abroad to pursue a PhD, showing patience over years of long-distance, and meeting my endless desire for Shoryu ramen everytime you visited.

To my Mum and Dad. You are always there in every step I take, you gave me the opportunity to get to know myself and what I was passionate of. Your love and care supported me throughout these 29 years. To my brother Giovanni, you have come a long way, inspiring me to work hard and aspire to great things.

All of you contributed to my growth and my path.

Gea per primo generò, simile a sé,

Urano stellato, che l'avvolgesse tutta d'intorno

— Esiodo, Teogonia

Thesis roadmap

Despite being robust over a wide range of observables, the standard cosmological model is facing open fundamental questions, such as those related to the physical nature of dark energy and dark matter. Given a model, this can be tested against observations by performing parameter inference and the larger and more accurate the datasets will be the better the constraints. However, the increase in data sample size and model complexity has established numerical challenges that traditional inference methods can hardly face. This urged the need for data compression and simulation-based frameworks, which are at the base of the work here presented.

This thesis starts in Part I by providing an introduction to both the standard cosmological model and cosmological inference, setting some background knowledge needed to understand the original work later presented. Chapter 1 provides an overview of the standard model, from the dynamics of the homogeneous Universe to a high-level introduction to structure formation. Chapter 2 expands on the probes of interest used to place constraints on cosmological parameters, namely gravitational waves and the Lyman-α forest. Finally, the standard inference framework is presented in Chapter 3, together with alternative approaches relevant to this work.

The main projects I have been working on during my PhD are presented in Part II, mainly split into two blocks. The first one refers to the work presented in Chapter 4 and it is built around testing the reliability of the simulation-based inference algorithm. In particular, we tested it while performing population-level inference of the Hubble constant from light standard sirens, simulated from a simple toy model. Given the latter, we are able to run both the traditional analytical and the likelihood-free approaches, testing the second against biases, with a particular focus on selection effects. The second block is linked to the study of how cosmological information can be extracted from Lyman-α forest correlation functions, alternatively to standard analyses, and it is developed in Chapter 5 and Chapter 6. First, in Chapter 5, we test the cosmological information contained in the full correlation functions, beyond the baryon acoustic oscillations peak, by extending the traditional framework to allow for direct cosmological inference. Then, in Chap-

ter 6 we apply score compression on realistic Lyα correlation functions and explore some of its outcomes, including reliable covariance matrix estimation and goodness of fit tests.

Finally, Part III summarises the main findings of these works and develops an outlook on a few extensions to the frameworks built here.

Contents

List of Figures

List of Tables

Part I

Introduction

1 The standard model of cosmology

Over the last century, our understanding of the evolution of the Universe throughout cosmic eras significantly progressed, culminating in the formulation of a consensus model, namely the ΛCDM paradigm. The latter is named after the two most abundant components in the Universe at present time, namely the cosmological constant Λ, responsible for the cosmic acceleration, and cold dark matter (CDM), responsible for structure formation. Despite the success of the model in explaining cosmological observations, there are still deep open questions we have not yet solved, such as the nature of the dominant dark energy and dark matter components. To further improve our comprehension and constrain our models, a good amount of diversified data and reliable inference tools are both indispensable. Testing standard inference analyses in cosmology and extensions to them is the goal of this work. The first part of this thesis provides an overview of the standard cosmological model in Chapter 1, further focusing on some probes of interest for this work in Chapter 2 and finally introducing the standard inference framework and some alternative approaches in Chapter 3.

The current modelling of the history and dynamics of the Universe is based on the Cosmological Principle, which assumes the Universe to be isotropic and homogeneous, at large scales. In other words, the Universe appears to be the same in every direction and from every location, respectively. However, we do have observational evidence that at small scales there is a significant deviation from homogeneity, and instead the Universe is highly structured.
This Chapter is structured as follows. In Sect. 1.1 I summarise the basics of the ΛCDM model, with a focus on the Universe components and dynamics, and their interplay. Then, in Sect. 1.2 I introduce inhomogeneities and only highlight the key concepts of structure formation that are propaedeutic for understanding part of the original work presented in Part II.

1.1 The homogeneous Universe

1.1.1 General Relativity framework

Einstein's theory of General Relativity (GR) (Einstein, 1915) has proven multiple times to be a robust mathematical description of the Universe's geometry and dynamics. Besides effectively explaining Solar System phenomena, such as the perihelion precession of Mercury's orbit, this theory confirmed its outstanding success with the first gravitational wave detection in 2015 (Abbott et al., 2016), a century later. This theory completely overturned the concept of gravity as a force, instead describing it as a geometrical property of the four-dimensional spacetime. In particular, the geometry of such a manifold is intrinsically connected to any energy source distribution as formally described by Einstein's field equations

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu} \tag{1.1}$$

$$\text{with } G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \, . \tag{1.2}$$

Throughout this introductory chapter we will assume the $c = 1$ convention for simplicity. Eq. (1.1) elegantly connects the Einstein tensor $G_{\mu\nu}$ on the left-hand side (LHS), which incorporates the spacetime's geometry and curvature, to the energy contribution $T_{\mu\nu}$ on the right-hand side (RHS), multiplied by Newton's gravitational constant G , and the cosmological constant Λ.

The Einstein tensor is a function of a metric tensor $g_{\mu\nu}$, which encodes the geometrical properties of the spacetime, and both the Ricci tensor $R_{\mu\nu}$ and scalar R, which relate to the local curvature induced by the source of $T_{\mu\nu}$. In particular, the concept of metric tensor $g_{\mu\nu}$ is at the core of the GR formalism. In an arbitrary coordinate system described by some coordinate $\mathbf{X} = (t, \mathbf{x})$, such a tensor returns the invariant distance between two infinitesimally close points via

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu \, , \tag{1.3}$$

where we used the Einstein summation convention.

In an isotropic and homogeneous Universe, $g_{\mu\nu}$ is given by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, such that the line element in polar spatial coordinates (r, θ, ϕ) is given by

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right] . \tag{1.4}$$

In this equation the dimensionless scale factor $a(t)$ accounts for the Universe expansion / contraction, where by convention $a(t_0) = 1$ at present time t_0, whereas k is the curvature constant, which can either be negative if the Universe is open, positive if closed and null if flat (Euclidean).

1.1.2 Background dynamics

Assuming a perfect fluid behaviour for the constituents of the Universe, the energy-momentum tensor can be written as $T_{\mu\nu} = diag(-\rho, P, P, P)$, where ρ is the energy density and P the isotropic fluid pressure. It is worth noticing here that the tensor is diagonal because the Cosmological Principal assumption does not allow anisotropic stress. Given this expression for $T_{\mu\nu}$, in a FLRW spacetime, Einstein's equations lead to the following Friedmann equations

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} + \frac{\Lambda}{3} , \tag{1.5}$$

$$\frac{\ddot{a}}{a} = - \frac{4\pi G}{3} (\rho + 3P) + \frac{\Lambda}{3} , \tag{1.6}$$

where $\dot{a}$ and $\ddot{a}$ are the first and second time derivative of the scale factor. As the Universe changes its size with time, in principle either expanding or contracting, it is useful to define the Hubble and the deceleration parameters

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \qquad\qquad q(t) \equiv - \frac{a(t) \ddot{a}(t)}{\dot{a}(t)^2} , \tag{1.7}$$

which encode the rate of expansion/contraction and how quickly this same rate varies, respectively. We refer to their value at present time t_0 as the Hubble constant H_0, which is one of the standard parameters of the cosmological model, and q_0.

The conservation of the energy-momentum $\nabla_\mu T^\mu_\mu = 0$, where ∇_μ denotes the covariant derivative, yields

$$\dot{\rho} + 3H(\rho + P) = 0\,, \tag{1.8}$$

which combined to the Friedmann equations allows for the following solution

$$\begin{cases} a(t) \propto t^{\frac{2}{3(1+P/\rho)}} & \text{if } \; P \neq -\rho \\ a(t) \propto e^{Ht} & \text{otherwise} \end{cases}\,, \tag{1.9}$$

which explicitly relates how the scale factor changes as a function of the properties of a given fluid.

1.1.3 Distances

Because of the cosmic flow or local peculiar motions, the objects we see can move either further or closer to the observer. As they move, the spectra coming from a light source will present a shift in observed wavelength (λ_o) with respect to the one (λ_e) at emission time t_e, such that we can define the redshift

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e}. \tag{1.10}$$

We refer to the cosmological redshift as that sourced by the Universe background dynamics, and it is related to the scale factor via

$$1 + z = \frac{1}{a(t_e)}. \tag{1.11}$$

In 1929, Hubble provided experimental evidence of the Universe expansion, reconstructing the diagram presented in Fig. 1.1, which shows how the target extra-galactic nebulae were moving further from the observer proportionally to their distance. This is a striking example of how distances are an incredibly powerful tool to constrain the expansion history, leading decades later to the discovery of the cosmic acceleration by Riess et al. (1998) and Perlmutter et al. (1999), who used luminosity distances of Type Ia supernovae. In cosmology we distinguish between *luminosity* and *angular diameter distances*, which are respectively derived from the flux and geometry of

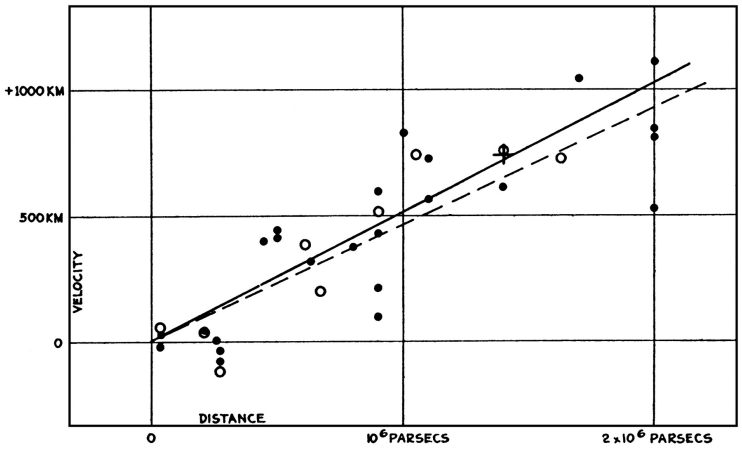

Figure 1.1: Original figure taken from Hubble (1929), it plots the velocity–distance relation among extra-galactic nebulae.

objects.

Let us assume flatness for simplicity and let us first derive an expression for the luminosity distance via a simple reasoning. The flux is related to the luminosity of a source via the distance of the latter by

$$F = \frac{L}{4\pi D_{\rm L}^2} \,, \tag{1.12}$$

being $D_{\rm L}$ the newly defined luminosity distance. Because of the expansion, the loss of energy of photons along the path and the difference between λ_e and λ_o needs to be taken into account, leading to an effective form for the flux

$$F = \frac{L}{4\pi D_{\rm M}^2 (1+z)^2} \,, \tag{1.13}$$

being $D_{\rm M}$ the comoving distance between the observer and the light source, defined as

$$D_{\rm M} = \int_0^z \frac{dz}{H(z)}. \tag{1.14}$$

By definition of $D_{\rm M}$, then the luminosity distance will be given by

$$D_{\rm L} = (1+z)\int_0^z \frac{dz}{H(z)}. \tag{1.15}$$

Rather than the flux of a source, we might instead be interested in using its geometrical properties. Under the assumption of a flat FLRW Universe, in the limit of small angles, for a distant object of fixed physical size l and angular diameter $d\theta$ on the sky, the distance to the object is $D_{\rm A} = \frac{l}{d\theta}$. Still under the hypothesis of a flat Universe, the angle $d\theta$ is equal to the comoving size of the object l/a divided by the comoving distance $D_{\rm M}$, which yields

$$D_{\rm A} = \frac{D_{\rm M}}{(1+z)}, \tag{1.16}$$

where we have used Eq. (1.11) to express $D_{\rm A}$ as a function of redshift.

1.1.4 Energy content

As already seen with the Friedmann equations, under perfect isotropic fluid assumptions, the constituents of the Universe contribute to the rate of expansion with their energy density. How much all these components contribute to the total energy budget changes over time, given that, by integration of Eq. (1.8), each of them has a density that varies with the scale factor as

$$\rho \propto a^{-3(1+w)}\,, \tag{1.17}$$

being $w = P/\rho$ the equation of state for that component. Defining the critical density today as $\rho_c = \dfrac{3H_0^2}{8\pi G}$ (in this notation $\rho_c = \rho_{c,0}$), the energy contribution of a certain constituent i is usually expressed in terms of the density parameter at present time

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} = \frac{8\pi G}{3H_0^2}\rho_i\,. \tag{1.18}$$

The energy budget components are:

- ***Baryons***. By baryons we indicate all ordinary matter that interacts both gravitationally and electromagnetically, which energy density varies as $\rho_b \propto a^{-3}$, with their equation of state $w_b = 0$. At present time baryons only account for approximately 5% of the total energy budget, taking either the form of objects, like stars and galaxies, or diffuse gas. At early times, baryons were characterised by a tight coupling to photons, but as the primordial plasma reached sufficiently low temperatures, the first atoms started forming, defining the epoch of *recombination*. Subsequently, the number of free electrons dropped and baryons effectively *decoupled* from radiation. Hydrogen atoms became neutral and baryonic particles were free to collapse to form structures; further details will be given in Sect. 1.2.

- ***Dark Matter***. By dark matter we indicate all matter that behaves like a pressureless fluid ($w_c = 0$) but does not interact electromagnetically, hence 'dark'. At present time, it accounts for almost 27% of the total energy budget. First observed via galaxies' rotation curves (Rubin and Ford, 1970), its particle nature has not been yet understood, however we do have experimental evidence from structure formation that it is 'cold' dark matter, namely characterised by non-relativistic velocities. Indeed, if its thermal velocity would have been too large, only structures located in larger potential wells would have formed, which is not what we observe.

- ***Dark Energy***. Today it is the most abundant component, with approximately $\Omega_\Lambda \approx 0.7$, and it generically refers to the mechanism which is driving the current acceleration phase of the Universe expansion, only recently discovered in 1998 (Perlmutter et al., 1999; Riess et al., 1998). The nature of this component is unclear and there is a plethora of currently explored models, among which the current standard model prefers the cosmological constant Λ. The latter is characterised by an equation of state equal to $w_\Lambda = -1$, such that its energy density is constant over time.

- ***Photons***. Relativistic particles characterised by $w_\gamma = 1/3$, and hence energy density that varies as $\rho_\gamma \propto a^{-4}$. Most of the photons we currently see are relics from the Big Bang and

constitute the cosmic microwave background (CMB). Despite their small contribution at present time, with $\Omega_r \approx 10^{-5}$, at early times they were the dominant source of energy.

- ***Neutrinos***. These particles are not only characterised by weak and gravitational interactions, but they also oscillate between different mass states. Understanding the nature of neutrinos and placing a constraint on the sum of their masses are active areas of research. Initially neutrinos were relativistic and part of the *radiation* component, turning into a non-relativistic pressureless species around $z \sim 100$.

- ***Curvature***. In case of non-zero curvature, this can be included among the components to the energy budget, characterised by $w_k = -1/3$, as per Eq. (1.5). Current observations strongly favour a flat Universe (Planck Collaboration et al., 2020; Alam et al., 2021).

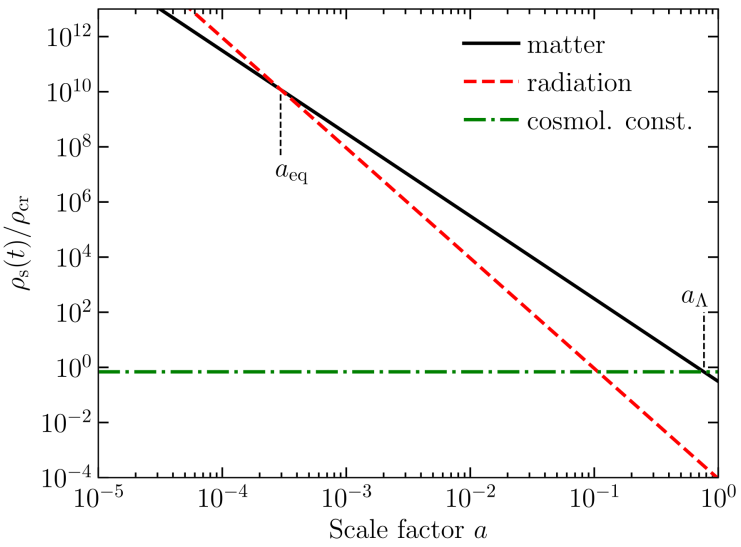

Figure 1.2: This figure represents the evolution of components' energy densities over time. The two epochs referred to $a_{\rm eq}$ and a_Λ are the epochs at which the matter energy density equals the radiation and cosmological constant ones, respectively. Image taken from Dodelson and Schmidt (2020).

Grouping all the above components into three main classes of fluids, namely *radiation* ($w_r = 1/3$), *matter* ($w_m = 0$) and the cosmological constant ($w_\Lambda = -1$), and expressing the curvature contribution in terms of $\Omega_k = -k/H_0^2$, the Friedmann equation in Eq. (1.5) can be rewritten as a function of the density parameters at present time as

$$H^2 = H_0^2(\Omega_{\rm r} a^{-4} + \Omega_{\rm m} a^{-3} + \Omega_{\rm k} a^{-2} + \Omega_\Lambda)\,. \tag{1.19}$$

The expansion rate is a direct effect of which component is dominant during a certain cosmological epoch. From early to late times, it is possible to distinguish, as clearly visible in Fig. 1.2, the radiation, matter and dark energy domination eras.

1.2 Inhomogeneities

Despite being homogeneous and isotropic at large scales, the Universe is highly structured at smaller scales. The presence of matter overdensities, in the form of filaments and nodes, over voids is visible in the *cosmic web* representation in Fig. 1.3. These deviations from the homogene-

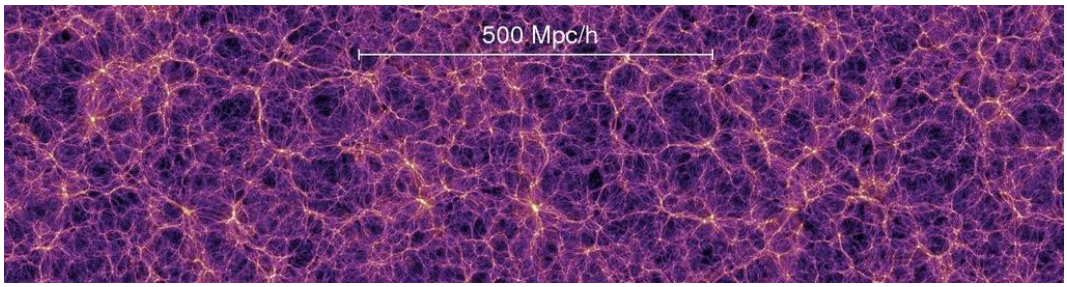

Figure 1.3: Snapshot of the Universe at present time, from the Millennium Simulations, available at https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/.

ity discussed in Sect. 1.1 can be statistically characterised as a function of the cosmological scale. To this aim, in cosmology we usually refer to either a two-point correlation function in real space or, equivalently, to its Fourier transform, the *power spectrum*. For a given fluid i, at coordinate $\mathbf{x}$ and time t, it is possible to define the overdensity field $\delta_i(\mathbf{x}, t)$ with respect to its mean density $\bar{\rho}_i(t)$ as

$$\delta_i(\mathbf{x}, t) = \frac{\rho_i(\mathbf{x}, t) - \bar{\rho}_i(t)}{\bar{\rho}_i(t)} \tag{1.20}$$

By definition of Eq. (1.20), the mean value of the overdensity field must be zero, $\langle \delta(\mathbf{x}, t) \rangle = 0$. The Fourier transform of the $\delta_i(\mathbf{x}, t)$ can be written as

$$\delta_i(\mathbf{k}, t) = \int \delta_i(\mathbf{x}, t) e^{-i\mathbf{k}\cdot\mathbf{x}} d^3\mathbf{x} \,, \tag{1.21}$$

with $\mathbf{k}$ the 3D wavevector, which amplitude will be the wavenumber $k = |\mathbf{k}|$. Considering two regions of space, separated by an infinitesimal $\mathbf{r}$, at the same time t, we define the two-point correlation function as

$$\xi(|\mathbf{r}|, t) = \langle \delta(\mathbf{x}, t), \delta(\mathbf{x} + \mathbf{r}, t) \rangle \tag{1.22}$$

and the power spectrum $P(k, t)$ to be its Fourier transform, which satisfies

$$\langle \delta(\mathbf{k}, t) \delta(\mathbf{k}', t) \rangle = (2\pi)^3 P(k, t) \delta^D(\mathbf{k} + \mathbf{k}'). \tag{1.23}$$

The time evolution of the power spectrum is of high interest for cosmology. According to the standard cosmological model, the seeds to primordial perturbations were initially sourced by an inflationary mechanism, an early rapid expansion phase (Guth, 1981; Starobinsky, 1980). Inflation also generates tensor fluctuations, but that is not of interest for this work, as we will see in what follows.

1.2.1 Scalar perturbations

In GR we distinguish between scalar, vectorial and tensorial perturbations. At linear order, they evolve independently (Dodelson and Schmidt, 2020), which means they can be conveniently studied separately. We are restricting the analysis to scalar perturbations only, as they are the most relevant for structure formation. While vector perturbations have only decaying modes, tensor fluctuations are instead linked to the generation of gravitational waves, beyond the scope of this introduction, and they do not couple to the matter and radiation densities, not being responsible

for the large-scale structure. Finally, in cosmology, we assume these perturbations to be adiabatic, as predicted by simple inflationary models.

The evolution of perturbations with respect to the underlying homogeneous background can be studied by perturbing the quantities that enter the Einstein field equations. In particular, any perturbed quantity x assumes the form of $x = \bar{x} + \delta x$, where the two terms are, respectively, the unperturbed background value and the perturbation, which is assumed to be small, such that $\delta x \ll \bar{x}$. Whenever this condition is not satisfied we enter the non-linear regime.

Starting from the LHS of Eq. (1.1), the basic component to perturb is the FLRW metric, assuming flatness for simplicity. This poses a problem of coordinate choice, also known as *gauge choice*, which is non-trivial. Some choices can drastically simplify the calculations; we restrict ourselves to the Newtonian gauge so that the perturbed line element becomes

$$ds^2 = a^2(\tau)[-(1+2\Psi)d\tau^2 + (1-2\Phi)\delta_{ij}dx^i dx^j], \tag{1.24}$$

where Ψ and Φ are the perturbation variables and $\tau = \int_0^t \frac{t}{a(t)} dt$ is the conformal time. The two variables Ψ and Φ can be respectively thought as the Newtonian potential and the perturbation to the spatial curvature. On the other hand, perturbations to the RHS of the Einstein field equations correspond to perturbing the energy-momentum tensor, such that $T^\mu_\nu = \bar{T}^\mu_\nu + \delta T^\mu_\nu$. This could possibly introduce a form of anisotropic stress, but we require that to be zero, so that $\Psi = \Phi$.

The evolution of perturbative modes $\delta_i(\boldsymbol{k})$ (see Eq. (1.21)) primarily depends on their scale. All modes start as *super-horizon*, meaning their wavenumber k is smaller than the comoving Hubble radius $\mathcal{H} \equiv aH$, and as the Universe evolves and the horizon expands, modes progressively cross it, becoming *sub-horizon*. Effectively the horizon serves as a threshold between two different regimes, as the perturbations evolve differently depending on whether their scale is larger or lower than it. Secondly, their evolution is also affected by the cosmological era, namely by the dominant

source of energy at a certain epoch, either radiation, matter or dark energy (late Universe). As an example, this is evident from the equation for the evolution of the perturbation Φ

$$\Phi'' + 3(1+w)\mathcal{H}\Phi' + wk^2\Phi = 0\,, \tag{1.25}$$

where $'$ denotes derivatives with respect to conformal time τ and $\mathcal{H} \equiv aH$. The dependence on the background dynamics, or equivalently on the cosmological era, is expressed via the equation of state w, and the Hubble parameter, and also the third addend on the LHS depends on the scale. The evolution of perturbations through the radiation and matter epochs can be encoded into the *transfer function* $T(k)$, which help us relate the perturbations we observe today to the primordial curvature perturbations $\mathcal{R}(\mathbf{k})$ via (Dodelson and Schmidt, 2020)

$$\Phi(\mathbf{k}, a) \sim \mathcal{R}(\mathbf{k}) \times \{\text{Transfer function(k)}\} \times \{\text{Growth factor(a)}\}, \tag{1.26}$$

where the growth factor accounts for the amplitude growth at late times, well within the matter domination ($a_{\rm late}$), such that, for $a > a_{\rm late}$,

$$\frac{\Phi(\mathbf{k}, a)}{\Phi(\mathbf{k}, a_{\rm late})} \equiv \frac{D_+(a)}{a}\,. \tag{1.27}$$

We also here define a related quantity that will be useful throughout the thesis, the *growth rate*

$$f(a) \equiv \frac{d\ln D_+(a)}{d\ln a}, \tag{1.28}$$

which is related to the matter density via $f(a) \simeq \Omega_m(a)^{0.55}$ (Peebles, 1980).

The primordial power spectrum for curvature perturbations $\mathcal{R}$ as predicted by inflation is

$$P_{\mathcal{R}}(k) = 2\pi^2 k^{-3} A_s \left(\frac{k}{k_P}\right)^{n_s - 1}\,, \tag{1.29}$$

where A_s and n_s are respectively the amplitude of scalar perturbations and the scalar index, while k_P is some pivot scale. Finally, given the decomposition in Eq. (1.26), the linear power spectrum of matter at late times can be obtained from the $P_{\mathcal{R}}(k)$ as (Dodelson and Schmidt, 2020)

$$P_{\rm lin}(k,a) \propto D_+^2(a) T^2(k) P_{\mathcal{R}}(k) \,. \tag{1.30}$$

Evolution of single fluids

Considering now a single fluid i, the evolution for its density contrast $\delta_i = \frac{\delta\rho_i}{\rho_i}$ is obtained by a combination of the mass and momentum conservation on sub-horizon scales, in Fourier space, and the Poisson equation. The first two equations, which are the continuity and Euler equations, are respectively obtained from the $\nu = 0$ and $\nu = i$ components of the perturbed stress-energy conservation law

$$\nabla_\mu T^\mu_\nu = 0 \,. \tag{1.31}$$

Instead, the Poisson equation is obtained by combining the 00 and $0i$ components of the Einstein field equations. The resulting equation for the evolution of a fluid perturbation is

$$\ddot{\delta}_i + 2H\dot{\delta}_i + \frac{c_s^2}{a^2}\left(k^2 - k_J^2\right)\delta_i = 0 \,, \tag{1.32}$$

where $c_s^2 = d\bar{P}_i/d\bar{\rho}_i$ is the speed of sound and $k_J = \sqrt{4\pi G\bar{\rho}a^2/c_s^2}$ is the Jeans scale (Jeans, 1902). Eq. (1.32) takes into account three different effects: Hubble drag, fluid pressure and gravity. Not only the background dynamics affect the evolution of perturbation modes, as earlier mentioned, but there is also a fine interplay between gravitational collapse and pressure effects. On small scales, characterised by $k \gg k_J$, the solution to Eq. (1.32) is a damped oscillator: while pressure drives the oscillation opposing to the gravitational collapse, the Hubble friction damps it. On the other hand, at large scales ($k \ll k_J$) the solutions are growing modes.

Let us now consider the evolution of dark matter perturbations through the epoch of matter-radiation equality (t_{eq}), namely when the energy densities of matter and radiation were equal.

Once defined a new variable $y = a/a(t_{eq})$, it is possible to rearrange Eq. (1.32) for the evolution of cold dark matter perturbations δ_c, for which $w_c = 0$ and c_s can be ignored, into the Mészáros equation (Meszaros, 1974)

$$\frac{d^2\delta_c}{dy^2} + \frac{2+3y}{2y(y+1)}\frac{d\delta_c}{dy} - \frac{3}{2y(y+1)}\delta_c = 0\,. \tag{1.33}$$

This equation admits two separate solutions:

$$\delta_c \propto \begin{cases} 2+3y \\ (2+3y)\ln\left[\dfrac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right] - 6\sqrt{1+y} \end{cases}\,. \tag{1.34}$$

During radiation domination ($y \ll 1$), on most scales the ('dominant') solution is $\delta_c \propto \ln a$: the rate at which dark matter collapse is slowed down by the effect of radiation. On the other hand, for $y \gg 1$, it grows faster as the first solution to Eq. (1.33) is $\delta_c \propto a$, while the second falls off as $y^{-3/2}$. Finally in the late Universe, when dark energy starts dominating, the growth is suppressed and instead $\delta_c = const$ on all scales.

Photon-baryon oscillations

At earlier times, before the epoch of recombination, baryons and photons were tightly coupled due to Compton scattering, effectively behaving as a single fluid. The evolution of the baryons and photons density contrasts, which are respectively denoted as δ_b and δ_γ, follow (Peebles and Yu, 1970; Hu and White, 1996; Eisenstein et al., 2007)

$$\delta_b'' + \frac{\mathcal{H}R}{1+R}\delta_b' - c_s^2\nabla^2\delta_b = \nabla^2\Phi = \frac{3}{4}\left(\delta_\gamma'' + \frac{\mathcal{H}R}{1+R}\delta_\gamma' - c_s^2\nabla^2\delta_\gamma\right)\,, \tag{1.35}$$

where $R = (3\bar{\rho}_b)/(4\bar{\rho}_\gamma)$ and the speed of sound in the coupled fluid is $c_s = c/\sqrt{3(1+R)}$. Eq. (1.35) is the equation for a driven oscillator with frequency $c_s^2 k$: as the baryons would tend to collapse inside the potential well, photons provide pressure to support them, resulting in an oscillatory behaviour. However at the epoch of recombination, when the decoupling happens, the

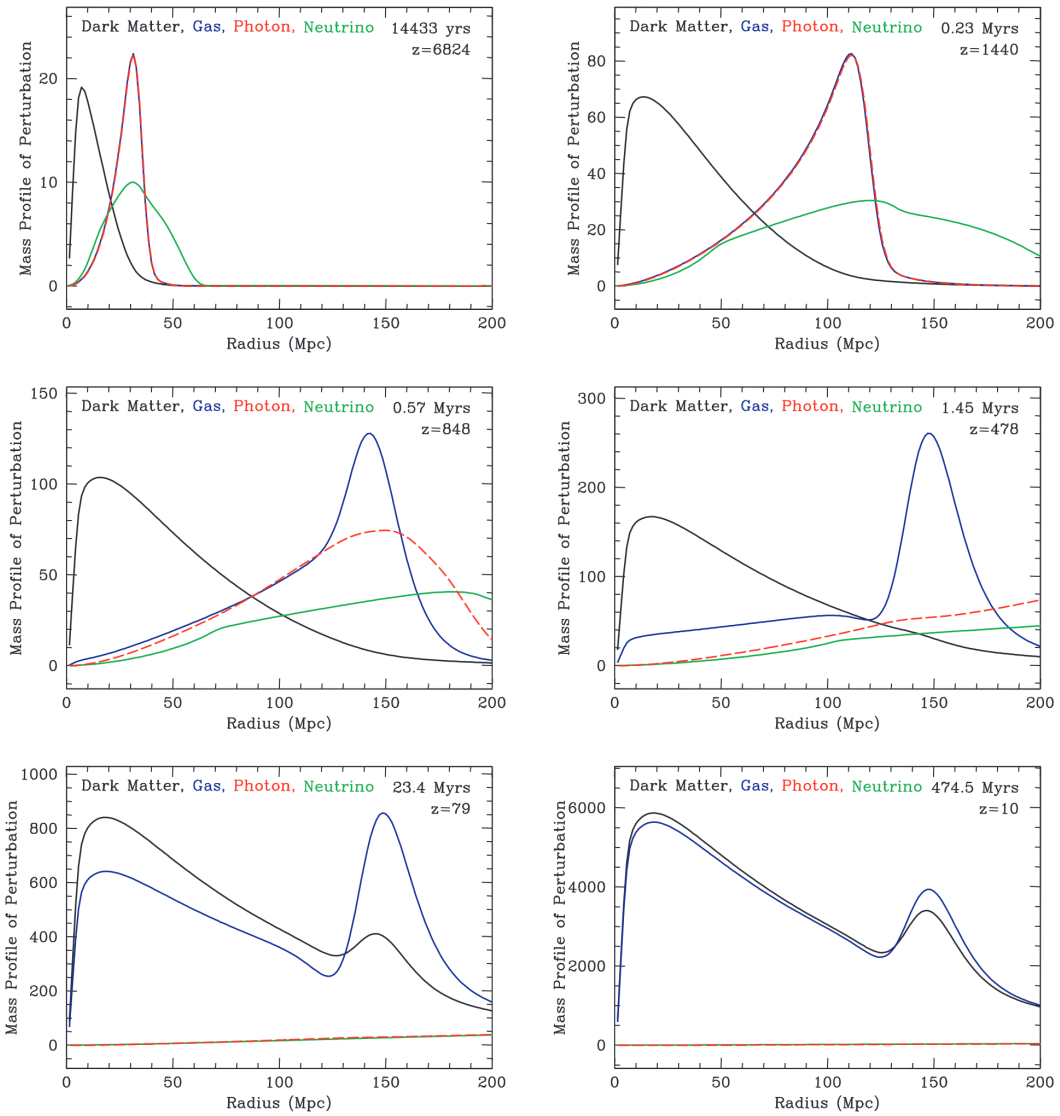

Figure 1.4: This plot shows the evolution over time (different panels correspond to different reshifts, starting from earlier times) of the perturbations mass profile for dark matter (black), baryons (blue), photons (red) and neutrinos (green), around a localised adiabatic perturbation. For a description refer to the text. Image taken from Eisenstein et al. (2007)

sustaining pressure disappears and the oscillations freeze leaving an imprint on both the distributions of baryons and photons, which we call *baryon acoustic oscillations* (BAO). The creation of the imprint at the level of the dark matter distribution is schematically shown in Fig. 1.4[1], where the frames are snapshots of the perturbation mass profile evolution over time for a single localised perturbation, as a function of the radius. While dark matter is left at the centre of the perturbation and neutrinos are able to free-stream, photons and baryons modes travel alongside up to a certain radius, which corresponds to the sound horizon at the drag epoch $z_d \simeq 1020$

$$r_d = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)} dz \,, \tag{1.36}$$

when baryons decoupled from photons, hence stopped being 'dragged' by them. After decoupling, photons are able to free-stream, and while baryons modes grow as $\delta_b \propto a$, like dark matter ones; these two components also gravitationally interact and fall into each other's overdensities, resulting in the configuration in the final panel of Fig. 1.4. In the real Universe, many of these perturbations overlap, which results in an increased clustering at the typical comoving radius $r_d \simeq 150\mathrm{Mpc}$ from an overdensity. This emerges as a peak in matter tracer's correlation functions, as we will later see.

1.2.2 The effect of peculiar velocities

As part of the growth of structures there are peculiar velocities: in a local system, objects move with respect to the Hubble flow because of local interactions. These motions contribute to the redshift of the object, shifting it towards larger values if the source is moving away from us along the line of sight, or smaller if moving towards the observer. This results in the so called *redshift space distortions* (RSD), which are distortions in the matter distributions in redshift-space. Referring to the redshift-space instead of the real-space effectively means mapping the objects' position as a function of their redshift rather than their physical distance. In particular, for an object located at $\boldsymbol{x}$ in real-space, the position $\boldsymbol{s}$ in redshift-space will be equal to $\boldsymbol{s} = \boldsymbol{x} - u_{\boldsymbol{z}}(\boldsymbol{x})\hat{\boldsymbol{z}}$, where the shift

[1] A clearer animation can be found at https://scholar.harvard.edu/deisenstein/book/export/html/28634

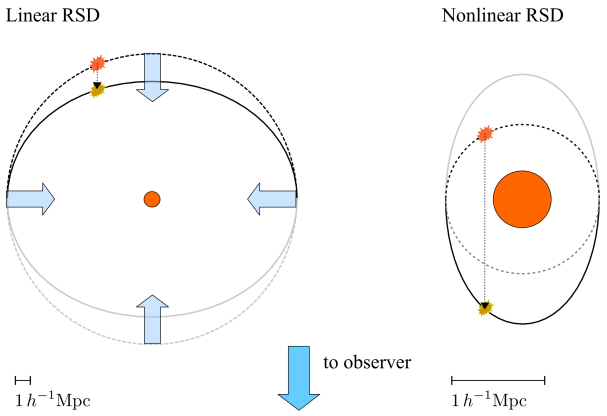

Figure 1.5: Redshift space distortions in both the linear (on the left) and non-linear (on the right) regimes, with respect to the central overdensity in orange, towards which the velocity flow (light blue arrows) is pointing: the dashed line refer to the distribution in real space, while solid in redshift space. The dotted arrow indicates the displacement caused by the line of sight velocity. Image taken from Dodelson and Schmidt (2020)

is induced by the peculiar velocity component $u_{\boldsymbol{z}}(\boldsymbol{x})$ along the line of sight axis $\hat{\boldsymbol{z}}$. The amplitude of the RSDs is set by the overall amplitude of the velocity power spectrum, which depends on the logarithmic growth rate f and on the linear power amplitude σ_8. The latter is the expected rms overdensity in a sphere of comoving radius $R = 8h^{-1}\mathrm{Mpc}$, defined as

$$\sigma_8^2 \equiv \langle \delta_{m,8}^2(\mathbf{x}) \rangle \tag{1.37}$$

$$\text{with } \delta_{m,R}(\mathbf{x}) \equiv \int \delta_m(\mathbf{x}') W_R(|\mathbf{x} - \mathbf{x}'|) d^3x' \,, \tag{1.38}$$

where $W_R(x)$ is a tophat window function with radius R.

At small scales, as it can be seen in Fig. 1.5, peculiar velocities contribute with non-linear redshift space distortions, causing elongations in redshift space, known as the *fingers of God*.

1.2.3 Beyond the linear regime

So far we restricted ourselves to the assumption of linear perturbations. However, at late times, as the modes keeps on evolving and matter collapses under gravity, highly non-linear structures form, in particular at smaller scales, such as galaxies. Studying the complex coupling between different

modes is non-trivial, and instead paved the way to the development of N-body and hydrodynamical simulations.

The non-linear evolution of structure also affects the BAO feature in the matter distribution: despite being more important at smaller scales, non-linearities also affect the larger scales and in particular cause a broadening of the acoustic peak. Eisenstein et al. (2007) demonstrated that this can be simply modelled in terms of a Gaussian smoothing with respect to the linear theory power spectrum, such that

$$P(\mathbf{k}) = P_{\rm lin}(\mathbf{k}) e^{\left(-\frac{k_\parallel^2}{2\sigma_\parallel^2} - \frac{k_\perp^2}{2\sigma_\perp^2}\right)} , \tag{1.39}$$

where $\mathbf{k} = \{k_\parallel, k_\perp\}$ is the wavenumber with components parallel and perpendicular to the line of sight and $(\sigma_\parallel, \sigma_\perp)$ are the corresponding smoothing parameters. These two parameters are predicted following the relations

$$\sigma_\perp = s_0 D \qquad\qquad \sigma_\parallel = s_0 D(1+f) , \tag{1.40}$$

being s_0 a cosmology-dependent length, D and f the growth factor and the growth rate, respectively.

2 Cosmological observables

In Chapter 1 we briefly introduced some basic building blocks of the ΛCDM paradigm, but cosmology is certainly not limited to it and includes a variety of theoretical models that also extend to more complex phenomena. However, the constraining power we have on those models depends on the amount of informative data we have from cosmological probes of interest. Over the last decades, we have collected a significant amount of data, over a wide range of observables, that enabled us to progressively tighten constraints on cosmological parameters. However, we are not yet at the point of clarifying for example the nature of dark energy and dark matter, and instead tensions between different datasets are arising, an example of which is a disagreement in the H_0 value (see Di Valentino et al. 2021 for a review). Usually the term 'H_0 *tension*' refers to the discrepancy, currently at 5.0σ, in the measurement of H_0 given by the Cosmic Microwave Background (CMB) (Planck Collaboration et al., 2020), $H_0 = (67.27 \pm 0.60)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and Type Ia supernovae calibrated on Cepheids (Riess et al., 2022), $H_0 = (73.04 \pm 1.04)\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. However, this disagreement can be further generalised distinguishing between two main groups of probes: early and late time probes. While the latter are direct model-independent measurements, the H_0 value inferred from early time probes is model-dependent, which means that potentially this tension could be a hint for physics beyond ΛCDM, if no systematics are found.

The CMB and Type Ia supernovae (Brout et al., 2022) are among the most famous constraining probes of ΛCDM, along with galaxy clustering (Alam et al., 2021) and weak lensing (Heymans et al., 2021; Abbott et al., 2023). However, in this Chapter, we will focus on the two I used throughout my work: gravitational waves and the Lyman-α forest. These only recently started to gain popularity and emerged as promising alternatives.

2.1 Gravitational waves

Gravitational waves (GW) are ripples of the spacetime manifold caused by highly energetic extreme phenomena that generate curvature fluctuations that indeed propagate like waves. Their existence was postulated already by Einstein in his theory of General Relativity and finally got empirically verified with the first detection in 2015 by LIGO (Abbott et al., 2016), paving the way to the series of observing runs for the current LIGO-Virgo-KAGRA Collaboration[1].

As already seen in Sect. 1.2, the equations for the evolution of perturbations are obtained by perturbing the terms in the Einstein field equations. Earlier we restricted ourselves to scalar components, whereas now we are interested in the tensor ones. A fully detailed mathematical derivation for these equations is outside the scope of this work and I will instead introduce the key quantities we are interested in. Under the assumption of a flat space, the perturbed metric can be written as an expansion around the Minkowskian metric $\eta_{\mu\nu}$

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \,, \tag{2.1}$$

in the limit of small perturbations ($|h_{\mu\nu}| \ll 1$). Assuming the wave propagates along the direction z with frequency ω and h_{ij} is symmetric and traceless, then (Hartle, 2003)

$$h_{ij}(t,z) = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos[\omega(t - z/c)] \,. \tag{2.2}$$

We define the two polarisations h_+ and $h_\times$ as *strains* of the wave.

Gravitational waves could either arise from primordial processes, single spinning massive objects or compact binary inspirals, but in the interest of this thesis I will only be considering GW originating from compact binaries. For a comprehensive review please refer to LSC-Virgo collabora-

[1] https://observing.docs.ligo.org/plan/

tion white paper and Riles (2013). By compact binaries we mean systems that are either made of

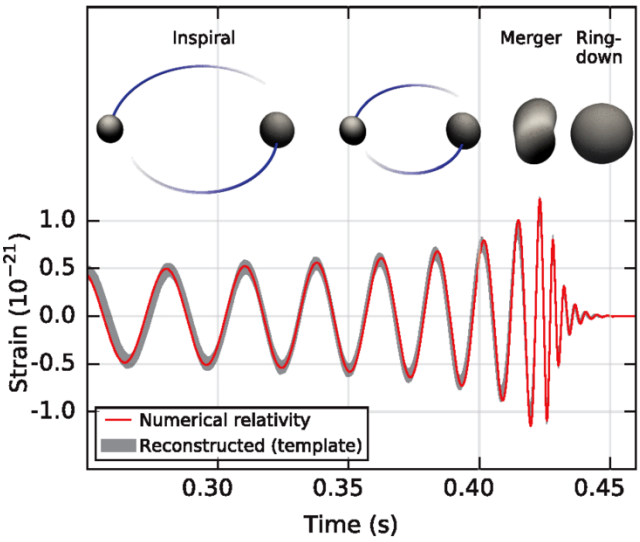

Figure 2.1: This plot shows the evolution of the detected strain as the system configuration changes. Figure adapted from Abbott et al. (2016)

binary neutron stars (BNS), binary black holes (BBH) or a NS-BH pair, that evolve as shown in the upper part of Fig. 2.1 through inspiral, merger and ringdown phases. As the objects start the inspiriling phase, they emit energy in the form of gravitational waves, while moving towards each other until merging. At that point, they coalesce into one object, characterised by the effective *chirp mass* M_c, reaching the highest strain amplitude, and finally the system settles down to its final state, during the ringdown phase.

Among the different science goals that can be achieved with these probes (see Sathyaprakash and Schutz 2009 and Bailes et al. 2021 for reviews), we are interested in placing a constraint on the Hubble constant H_0 (Mastrogiovanni et al., 2022), as further explained in Sect. 4.1. In particular, it was Schutz (1986) to first suggest using these systems to compute H_0, given both the localisation, hence an estimated redshift, and the fact that the amplitude of the fluctuation provides a self-calibrated measurement of their distance. This latter property is the reason why they are referred to as *standard sirens*. At time τ_{obs}, which is the time to coalescence as measured by the

observer, the average polarisation amplitude $h_c(\tau_{obs})$ for gravitational waves emitted by compact object mergers can be expressed as (Maggiore, 2007)

$$h_c(\tau_{obs}) = \frac{4}{D_{\rm L}}\left(\frac{G(1+z)M_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm gw}^{(\rm obs)}(\tau_{obs})}{c}\right)^{2/3}, \tag{2.3}$$

where $f_{\rm gw}^{(\rm obs)}(\tau_{obs})$ is the GW frequency measured by the observer and the dependence on the luminosity distance is highlighted in the first factor. For a Euclidean expanding Universe, knowing the redshift and the luminosity distance can lead to an estimate of the Hubble parameter via $H_0 = cz/D_{\rm L}(z)$, locally. The nature of the binary components hugely impacts the chance for an electromagnetic (EM) follow-up, and hence the precision of the H_0 constraint. If the EM counterpart is detected – *light* standard sirens – the host galaxy is identified and the redshift of the source can be measured, hence yielding H_0 when combined with D_L. Otherwise, if no EM counterpart is detected – *dark* standard sirens – the redshift can be estimated by summing up the contributions from all the possible host galaxies that are within the confidence region of the GW localisation Schutz (1986). Both these methods were tested on GW170817, which represents the very first detection of both electromagnetic and gravitational emission from a binary neutron star system. While Fishbach et al. (2019) applied the statistical approach, estimating $H_0 = 76.0^{+48.0}_{-23.0}$ km s^{-1}Mpc^{-1}, using the source redshift largely improved the constraining power, yielding $H_0 = 70.0^{+12.0}_{-8.0}$ km s^{-1}Mpc^{-1} (Abbott et al., 2017), see Fig. 2.2.

Motivated by the challenges of traditional Bayesian analyses for estimating H_0 from BNS systems, in Chapter 4 we test a simulation-based framework, where the likelihood function is not computed and instead estimated through neural networks given a set of simulations. The theory beyond this implementation will be first presented in Chapter 3.

2.2 Lyman-α forest

After being observed in the 1960s (Schmidt, 1965; Scheuer, 1965; Bahcall and Salpeter, 1965; Gunn and Peterson, 1965), the Lyman-α forest has recently proved its success as an alternative

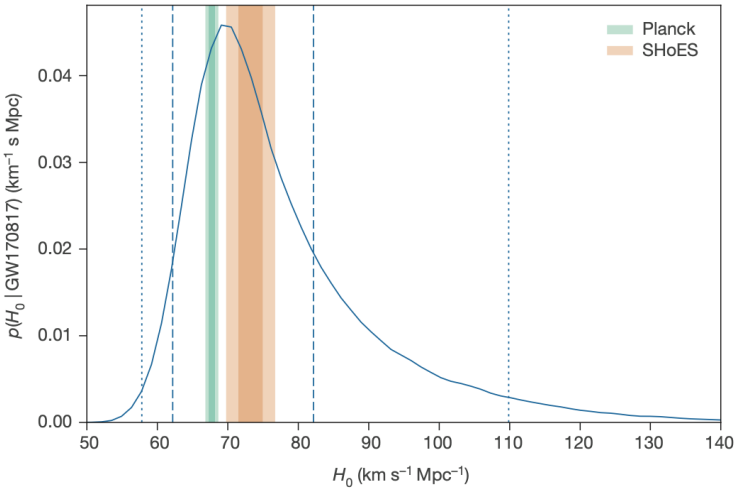

Figure 2.2: H_0 constraint from the GW170817 event (in blue) against the Planck (Planck Collaboration et al., 2016) and SH0ES (Riess et al., 2016) contours (in light green and orange respectively). Figure taken from Abbott et al. (2017).

probe for characterising matter distribution, probing the BAO peak at redshift $z > 2$ (Busca et al., 2013; Slosar et al., 2013; Kirkby et al., 2013; Font-Ribera et al., 2014). The Lyman-α forest, as shown in the lower panel of Fig. 2.3, is a series of absorption lines produced in high-redshift quasar (QSO) spectra by Lyman-α transitions of neutral hydrogen (HI) placed along the line the sight. The restframe wavelength at which the absorption happens is $\lambda_\alpha = 1215.67$Å . Those photons that are characterised at the time of emission by a wavelength $\lambda_e \leq \lambda_\alpha$ get redshifted as they travel and eventually, as they reach λ_α, they will undergo Lyα absorption if they are in a HI cloud. As a result, given the quasar and cloud redshifts $z_{\rm QSO}$ and $z_{\rm HI}$, absorption lines will be observed at $\lambda_{obs} = \lambda_e(1 + z_{\rm QSO}) = \lambda_\alpha(1 + z_{\rm HI})$, with a probability that depends on the properties of the neutral hydrogen in the inter-galactic medium (IGM).

The absorption probability is linked to the Lyα optical depth τ_α via $P \propto e^{-\tau_\alpha}$ and we are able to relate τ_α to gas with smooth line-of-sight velocity gradient dv/dx as per (McQuinn, 2016)

$$\tau_\alpha(z) = 1.3\delta_b\left(\frac{x_{\rm HI}}{10^{-5}}\right)\left(\frac{1+z}{4}\right)^{3/2}\left(\frac{H(z)}{1+z}\right)\left(\frac{dv}{dx}\right)^{-1}, \tag{2.4}$$

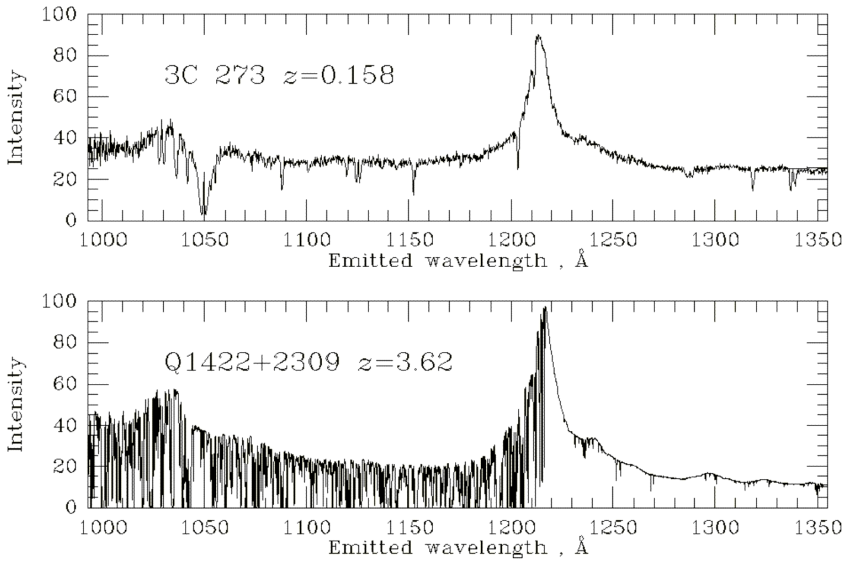

Figure 2.3: This figure shows the spectra as a function of the emitted wavelength for a low (upper panel) and high (lower panel) redshift QSOs, where the series of absorption lines is striking in the second case. Adapted from Bill Keel's website https://pages.astronomy.ua.edu/keel/agn/forest.html.

where δ_b is the baryon overdensity and $x_{\rm HI}$ the fraction of hydrogen that is neutral. As already mentioned in Sect. 1.2.2, peculiar velocities contribute with additional shifts in redshift, which here means that the final redshift attributed to a line will differ to the one caused by the Hubble flow alone. However, introducing gradients in the velocity field make things more complex, effectively changing the shape of the line and hence the value of the overdensities.

The observed quantity of interest is the transmitted flux and what we care about for the analysis are its correlation functions, given the flux fluctuations

$$\delta_F(\lambda) = \frac{f(\lambda)}{\bar{F}(\lambda)C(\lambda)} - 1 \tag{2.5}$$

as a function of the flux $f(\lambda)$, the mean transmitted flux fraction $\bar{F}(\lambda)$ and the continuum $C(\lambda)$, which are all related by $F = f/C$. The continuum of the spectrum is defined as the unabsorbed spectrum of the quasar, which we can only estimate.

In linear theory, define the cosine of the line-of-sight angle μ, the overdensity related to discrete matter tracers can be expressed in the form of

$$\delta^s_{\rm tracer}(k) = (b + \mu^2 f)\delta_{\rm m} \ , \tag{2.6}$$

which accounts for two main results of the growth of structures theory. First of all, discrete matter tracers, such as galaxies or quasars, are *biased* with respect to the dark matter distribution, which means that their particular distribution does not match the one of matter, as for example we know that galaxies only form where the density distribution peaks. This at linear order can be expressed as $\delta_{\rm tracer}(\mathbf{x}) = b_{\rm tracer}\delta_{\rm m}(\mathbf{x})$ via some linear bias parameter $b_{\rm tracer}$ (Bardeen et al., 1986; Dekel and Lahav, 1999; Sheth and Tormen, 1999). Secondly, tracers are not mapped in real-space, but rather in redshift-space, where peculiar velocities affect the position of the sources: if a tracer is located at $\boldsymbol{x}$, the position $\boldsymbol{s}$ in redshift-space will be shifted as in $\boldsymbol{s} = \boldsymbol{x} - u_{\boldsymbol{z}}(\boldsymbol{x})\hat{\boldsymbol{z}}$ by the peculiar velocity component $u_{\boldsymbol{z}}(\boldsymbol{x})$ along the line-of-sight axis $\hat{\boldsymbol{z}}$. Assuming an irrotational velocity field, the projection of the velocity along the line-of-sight can be expressed as $u_{\boldsymbol{z}} = \partial/\partial z \boldsymbol{\nabla}^{-2}\theta$, with $\theta \equiv \boldsymbol{\nabla}\cdot\boldsymbol{u}$ being the velocity divergence. In linear theory, the tracer's overdensity in redshift-space can be then obtained as (Percival and White, 2009)

$$\delta^s_{\rm tracer}(k) = \delta_{\rm tracer}(k) - \mu^2\theta(k) \ , \tag{2.7}$$

given the cosine of the line-of-sight angle μ. Further assuming linear theory for the velocity field, $\theta(k) = -f\delta(k)$, where f is the linear growth rate (see Eq. 1.28).

The tracer power spectrum then assumes the form of

$$P^s_{\rm tracer}(k,\mu) = b^2(1 + \mu^2\beta)^2 P_{\rm lin}(k) \ , \tag{2.8}$$

where $\beta \equiv f/b$ (Kaiser, 1987) and $P_{\rm lin}(k)$ is the isotropic linear matter power spectrum.

In the case of the Lyα forest, the situation is slightly more complicated: the transmitted flux frac-

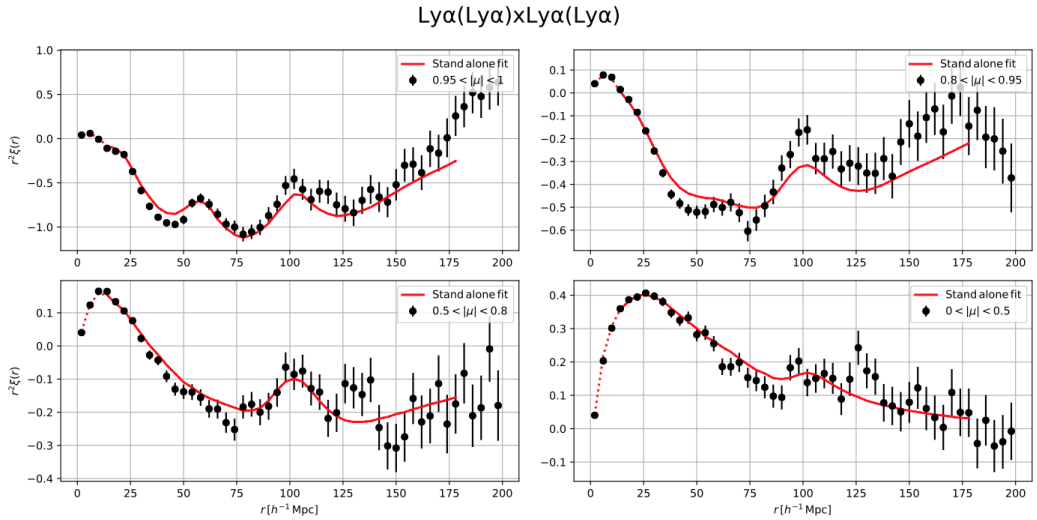

Figure 2.4: Plots for the 3D Lyα correlation function in bins of μ, together with the best-fit model. The BAO peak is visible at a separation of about $100h^{-1}$Mpc. Image taken from du Mas des Bourboux et al. (2020)

tion F is a non-linear function of the redshift-space optical depth and hence Eq. (2.6) needs to be modified (Seljak, 2012). This can be accounted for by introducing the Lyα velocity divergence bias factor $b_{\eta,\alpha}$, such that $\beta_\alpha \equiv \dfrac{b_{\eta,\alpha} f}{b_\alpha}$ (Arinyo-i-Prats et al., 2015), where b_α is the Lyα linear bias parameter. Therefore we can define the Lyα power spectrum as

$$P_\alpha(k,\mu) = b_\alpha(1+\beta_\alpha\mu^2)^2 P_{\rm lin}(k) = (b_\alpha + b_{\eta,\alpha}\mu^2 f)^2 P_{\rm lin}(k)\,. \tag{2.9}$$

Beyond this simple modelling, in reality it gets more complex and additional absorption can be caused by either heavier elements, generically defined as *metals*, and Damped Lyman alpha systems[2] (DLAs), which are characterised by a large column density of hydrogen, namely $\log(N_{HI}) > 20.3$. Metals and DLAs effectively act to contaminate the forest, and hence the correlation functions. DLAs produce complete absorption, over a wide range of wavelengths, surrounded by damped wings, which depend on the column density. These wings are modelled by a Voigt profile and can also affect the absorption line profiles at lower column densities, impacting the correlation function of the transmitted flux. Font-Ribera et al. (2012) showed that the broadening

[2] In the work that follows we will refer to high column density regions as HCDs.

effect of DLAs along the line-of-sight can be expressed in terms of new effective Lyα parameters via the definition of an extra function $F_{\rm DLA}$ which only depends on $k_{\parallel}$, since the absorption profiles of DLAs only manifest along the line-of-sight. The effective parameters will be defined as per (de Sainte Agathe et al., 2019)

$$b'_{\rm Ly\alpha} = b_{\rm Ly\alpha} + b_{\rm DLA} F_{\rm DLA}(k_{\parallel}) \,, \tag{2.10}$$

$$b'_{\rm Ly\alpha}\beta'_{\rm Ly\alpha} = b_{\rm Ly\alpha}\beta_{\rm Ly\alpha} + b_{\rm DLA}\beta_{\rm DLA} F_{\rm DLA}(k_{\parallel}) \,, \tag{2.11}$$

where $b_{\rm DLA}$ and $\beta_{\rm DLA}$ are the DLAs linear bias and RSD parameter, respectively. On the other hand, metals present in the intergalactic medium can also absorb the light emitted from the quasar at some discrete wavelenghts, within the Lyα forest region of the spectra, hence contaminating it. At the level of the Lyα correlation function, metal lines cause the appearance of extra bumps near the line-of-sight, at a separation that corresponds to the relative wavelength separation between the metal and the Lyα lines. As an example, in Fig. 2.4, additionally to the BAO feature, extra bumps are visible at $\sim 55-60{\rm Mpc/h}$, due to SiII(1190Å) and SiII(1193Å), $\sim 21{\rm Mpc/h}$, caused by SiIII(1207Å), and finally $\sim 105{\rm Mpc/h}$, less visible, as the effect of SiII(1260Å) (Farr et al., 2020). By defining for each metal line m_i a set of linear bias b_{m_i} and RSD parameter β_{m_i}, the contamination to the Lyα auto-correlation function $\xi_{\rm auto}$ will account for the correlation between each m_i and Lyα ($\xi_{{\rm Ly}\alpha\times m_i}$), and among all metal pairs ($\sum_{j\neq i}\xi_{m_i\times m_j}$). In cosmological analyses, metals and DLAs are treated as systematics.

The first 3D Lyα correlation measurement came from the Sloan Digital Sky Survey (SDSS) data (Slosar et al., 2011), with the most recent shown in Fig. 2.4, which uses SDSS DR16 data (du Mas des Bourboux et al., 2020). Current standard analyses of the Lyα correlation functions are used to measure the BAO feature, considering not only the Lyα auto-correlation function (Slosar et al., 2013; Busca et al., 2013; Kirkby et al., 2013), but also the cross-correlation of Lyα spectra with quasars position (Font-Ribera et al., 2014; du Mas des Bourboux et al., 2017; Blomqvist et al., 2019; du Mas des Bourboux et al., 2020). The availability of Lyα data will now

have an incredible boost with the data coming from the DESI (Dark Energy Spectroscopic Instrument) experiment, with the first EDR (DESI Collaboration et al., 2023) catalogue of Lyα forests already being published (Ramírez-Pérez et al., 2023). Such a new instrument will contribute with up to 1 million high-redshift quasars with $z > 2$ over five years of observations.

Current Lyα forests analyses are limited to infer cosmological information from the BAO peak component, this justifies the works presented in Chapter 5 and Chapter 6, where extensions to the current framework are explored and tested. In particular, the work presented in Chapter 5 aims at exploring whether there is extra information beyond the BAO peak alone coming from both the Lyα auto- and cross-correlation with quasars, also investigating which effects might contribute to it. Finally, in Chapter 6 we develop a framework for compressing the Lyα correlation data down and, as a secondary goal, testing the goodness of the estimated covariance matrix used in Lyα analyses. The approach we propose is *score compression*, which is a data compression algorithm I introduce in Sect. 3.2.1.

3 Data analysis techniques

Open questions in cosmology, such as the nature of dark energy and dark matter, have led to a plethora of theoretical models, among which it is difficult to prefer one over the other. On the other hand, over the last decades, there has been a boost in the instrumental power and amount of data we have to further probe the Universe dynamics and phenomena. In this *data-driven* era then it is crucial to set an efficient and reliable framework for the inference of model parameters given the data collected.

Most of the cosmological analyses rely on Bayesian statistics, which I introduce in Sect. 3.1, along with Monte Carlo algorithms for probabilistic computation. Despite being a well-tested framework, often its feasibility is challenged by the complexity of the tackled inference problems. Indeed, cosmological inference sometimes involves either too many dimensions or computationally expensive processes, which pave the way to alternative implementations, involving data compression and forward-modelling, respectively briefly discussed in Sect. 3.2. These concepts are at the basis of this thesis original work, outlined in Part II, where a more detailed methodological description is given.

3.1 Bayesian Inference

Probability $P(\mathcal{X}|\mathcal{M})$ is a mathematical property for an event $\mathcal{X}$ which defines how likely that event is to happen, provided a model $\mathcal{M}$. In particular, it is characterised by two fundamental equations called the *sum* and *product rules*, respectively

$$P(\mathcal{X}|\mathcal{M}) + P(\bar{\mathcal{X}}|\mathcal{M}) = 1\,, \tag{3.1}$$

$$P(\mathcal{X},\mathcal{Y}|\mathcal{M}) = P(\mathcal{X}|\mathcal{Y},\mathcal{M})P(\mathcal{Y}|\mathcal{M})\,, \tag{3.2}$$

where $\bar{\mathcal{X}}$ defines the event of $\mathcal{X}$ not happening and $\mathcal{Y}$ is another event. Because $P(\mathcal{X},\mathcal{Y}|\mathcal{M}) = P(\mathcal{Y},\mathcal{X}|\mathcal{M})$, Eq. (3.2) yields the Bayes theorem

$$P(\mathcal{X}|\mathcal{Y},\mathcal{M}) = \frac{P(\mathcal{Y}|\mathcal{X},\mathcal{M}) \times P(\mathcal{X}|\mathcal{M})}{P(\mathcal{Y}|\mathcal{M})}. \tag{3.3}$$

In a parameter inference context, if we assume $\mathcal{X}$ to be the model parameter θ and $\mathcal{Y}$ to be some set of data D, Eq. (3.3) can be rewritten as

$$P(\boldsymbol{\theta}|\boldsymbol{D},\mathcal{M}) = \frac{P(\boldsymbol{D}|\boldsymbol{\theta},\mathcal{M}) \times P(\boldsymbol{\theta}|\mathcal{M})}{P(\boldsymbol{D}|\mathcal{M})}\,, \tag{3.4}$$

which provides the probability of parameter values given the data, namely the *posterior distribution* $P(\boldsymbol{\theta}|\boldsymbol{D},\mathcal{M})$. In particular, the latter is proportional to the sampling distribution $P(\boldsymbol{D}|\boldsymbol{\theta},\mathcal{M})$, otherwise referred to as the *likelihood* function L when evaluated at the observed set of data, multiplied by the *prior* knowledge $P(\boldsymbol{\theta}|\mathcal{M})$ (also referred to as $\pi(\boldsymbol{\theta})$) about the parameters given the model $\mathcal{M}$. The denominator of Eq. (3.4) is called *evidence* $\mathcal{Z}$, and acts as a normalisation term for the posterior distribution.

Given a model described by N parameters, we might only be interested in learning about the probability distribution for a subset of them. To this aim, another fundamental concept for Bayesian statistics, obtained from Eqs. (3.1-3.2), is *marginalisation*. A simple demonstration for this principle is provided in Sivia and Skilling (2006). In the generic event $(\mathcal{X},\mathcal{Y})$ setup, where $\mathcal{Y}_i$ are all the possible values that $\mathcal{Y}$ can assume, marginalisation implies

$$P(\mathcal{X}|\mathcal{M}) = \sum_{\mathcal{Y}_i} P(\mathcal{X},\mathcal{Y}_i|\mathcal{M}), \tag{3.5}$$

becoming of the form

$$P(\theta_i|\boldsymbol{D},\mathcal{M}) = \int_{\theta_0} .. \int_{\theta_{i-1}} \int_{\theta_{i+1}} .. \int_{\theta_N} P(\hat{\boldsymbol{\theta}}|\boldsymbol{D},\mathcal{M}) d\hat{\boldsymbol{\theta}}, \tag{3.6}$$

in the continuous (data, parameters) space when we are solely interested in the posterior for the i-th parameter θ_i. Marginalisation is handy also for computing the evidence as

$$\mathcal{Z} \equiv P(\boldsymbol{D}|\mathcal{M}) = \int P(\boldsymbol{D}, \boldsymbol{\theta}|\mathcal{M}) d\boldsymbol{\theta} = \int P(\boldsymbol{D}|\boldsymbol{\theta}, \mathcal{M}) P(\boldsymbol{\theta}|\mathcal{M}) d\boldsymbol{\theta} \, . \tag{3.7}$$

3.1.1 Monte Carlo approaches

Monte Carlo methods are a class of algorithms based on repeated random sampling in order to estimate a certain target quantity. In the context of Bayesian statistics for cosmological analyses, probability distributions are approximated with the aid of samplers based on Markov chain Monte Carlo and nested sampling, introduced in the following paragraphs. I here only introduce these algorithms as their implementations are used within this thesis work, namely Cobaya for Metropolis-Hastings, PyStan for NUT sampling and finally Polychord for nested sampling. For a deeper investigation of these two classes of methods please refer to Gelman et al. (2013) and Skilling (2004) respectively.

Markov chain Monte Carlo methods

A Markov chain Monte Carlo (MCMC) is a memory-less random walk that draws values of the parameters set $\boldsymbol{\theta}$ in order to evaluate the target posterior distribution. Assuming that at the i-th step the algorithm starts from a point $\boldsymbol{\theta}^{i-1}$, first a $\boldsymbol{\theta}^*$ is randomly sampled from a proposal distribution and then the chain will eventually transition from $\boldsymbol{\theta}^{i-1}$ to $\boldsymbol{\theta}^*$, otherwise new proposals are generated. This walk is defined as memory-less because the probability of transitioning from $\boldsymbol{\theta}^{i-1}$ to $\boldsymbol{\theta}^*$ only depends on $\boldsymbol{\theta}^{i-1}$, and not on all the preceding samples. The requirement for the transition probability distribution $T(\boldsymbol{\theta}^i = \boldsymbol{\theta}^*|\boldsymbol{\theta}^{i-1})$ is to allow for convergence of the chain to the posterior distribution. I here briefly present two MCMC methods which have been used throughout the work: the classic Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) and Hamiltonian Monte Carlo (Neal, 2011; Betancourt, 2018).

i) Metropolis-Hastings algorithm The random walk begins by drawing a starting point $\boldsymbol{\theta}^0$ from a distribution $p_0(\boldsymbol{\theta})$ and sequentially (for $i = 1, 2, ..$ steps) sample proposal $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{i-1})$ until one new point is accepted and further proceed to the next step until convergence. Given the probability $p(\boldsymbol{\theta}|\boldsymbol{D})$ of drawing a generic set of parameters $\boldsymbol{\theta}$ conditioned on the data $\boldsymbol{D}$, the acceptance/rejection algorithm is based on the computation of the ratio

$$r = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{D})/q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{i-1})}{p(\boldsymbol{\theta}^{i-1}|\boldsymbol{D})/q(\boldsymbol{\theta}^{i-1}|\boldsymbol{\theta}^*)} \,, \tag{3.8}$$

such that

$$\boldsymbol{\theta}^i = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^{i-1}) = \min\{r, 1\} \\ \boldsymbol{\theta}^{i-1} & \text{otherwise} \end{cases} . \tag{3.9}$$

The transition probability is then given by $T(\boldsymbol{\theta}^i, \boldsymbol{\theta}^{i-1}) = \alpha(\boldsymbol{\theta}^i, \boldsymbol{\theta}^{i-1})q(\boldsymbol{\theta}^i, \boldsymbol{\theta}^{i-1})$. This ensures that the detailed balance

$$p(\boldsymbol{\theta}^i|\boldsymbol{D})T(\boldsymbol{\theta}^i, \boldsymbol{\theta}^{i-1}) = p(\boldsymbol{\theta}^{i-1}|\boldsymbol{D})T(\boldsymbol{\theta}^{i-1}, \boldsymbol{\theta}^i) \tag{3.10}$$

is fulfilled, sampled points will be correlated and $p(\boldsymbol{\theta}|\boldsymbol{D})$ is the equilibrium distribution of the Markov chain, to which the chain converges. Once the chain has converged, the algorithm recovers samples from the posterior distribution, finally yielding an estimate of the latter.

ii) Hamiltonian Monte Carlo (HMC) This is an MCMC algorithm that improves the efficiency of the random walk by introducing an auxiliary momenta variable $\boldsymbol{\phi}$ and building a Hamiltonian framework to run the sampling process. For each parameter θ_j there exists a momentum ϕ_j with an associated independent distribution $p(\boldsymbol{\phi})$, usually a multivariate normal distribution centred at 0 and with the covariance given by a 'mass matrix' M, and they are both sampled together in a way that $\boldsymbol{\phi}$ influences the transitioning rule of $\boldsymbol{\theta}$. In particular, given a first random draw of ϕ from $p(\phi)$, the sampling process follows a 'leapfrog steps' integration algorithm to update both θ^* and ϕ^*. This means that the update to $\boldsymbol{\theta}^*$ and

$\boldsymbol{\phi}^*$ happens after the following steps have been repeated a fixed number of times:

$$\begin{aligned} \boldsymbol{\phi} &\leftarrow \boldsymbol{\phi} + \frac{\epsilon}{2}\frac{\mathrm{d}\log p(\boldsymbol{\theta}|\boldsymbol{D})}{d\,\boldsymbol{\theta}} \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} + \epsilon M^{-1}\boldsymbol{\phi} \\ \boldsymbol{\phi} &\leftarrow \boldsymbol{\phi} + \frac{\epsilon}{2}\frac{\mathrm{d}\log p(\boldsymbol{\theta}|\boldsymbol{D})}{d\,\boldsymbol{\theta}} \end{aligned} \tag{3.11}$$

where ϵ is the discretization time in each leapfrog step. The acceptance/rejection step will follow as before with the acceptance ratio now being defined as

$$r = \frac{p(\boldsymbol{\theta}^*|\boldsymbol{D})p(\boldsymbol{\phi}^*)}{p(\boldsymbol{\theta}^{i-1}|\boldsymbol{D})p(\boldsymbol{\phi}^{i-1})}\,. \tag{3.12}$$

No U-turn sampler (NUTS). NUTS is an adaptive extension of HMC, where the number of leapfrog steps is not fixed a priori but is rather determined at each iteration in order to make the sampling more efficient (Hoffman and Gelman, 2011). In particular, the core idea is to avoid 'U-turns', namely resampling backwards along the same trajectory that has already been sampled (see Fig.3 of Neal 2011). To this end, the NUTS algorithm randomly samples either backwards or forwards in time a number of points which double at each leapfrog step (1 sample at step 1, 2 samples at step 2, 4 samples at step 3 and so on) – schematic two-dimensional representation in Fig. 3.1 – and adaptively stops the trajectory if the angle between the directions of two samples' gradients in a step is smaller than 90 degrees. At that point the sampling restarts in the same way from a point that has been sampled within the last set of consecutive steps in such a way that the trajectory will be different.

Nested sampling

Nested sampling is an alternative Monte Carlo approach which aims at estimating the evidence $\mathcal{Z}$. As the latter (Eq. 3.7) quickly becomes infeasible for a large parameter space, the idea behind this method (Skilling, 2006) is to transform it in the form of

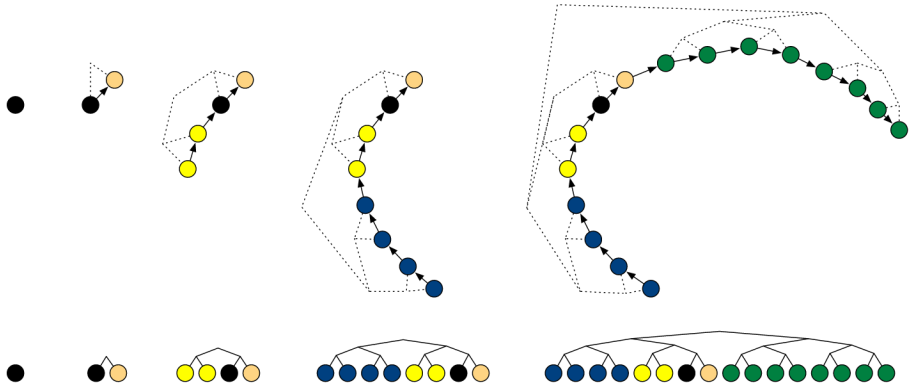

Figure 3.1: This figure shows an example of binary tree via subsequent doubling in a two-dimensional case. Image taken from Hoffman and Gelman (2011).

$$\mathcal{Z} = \int_0^1 L(X)dX \ . \tag{3.13}$$

X in Eq. (3.13) is the proportion of prior within the contour defined by $L = \lambda$:

$$X(\lambda) = \int\int .. \int_{L(\boldsymbol{\theta})>\lambda} \pi(\boldsymbol{\theta})d\boldsymbol{\theta} \ , \tag{3.14}$$

and it is a decreasing function of the likelihood (see Fig. 3.2). The sampling algorithm first starts by randomly sampling N live points from the full prior and proceeds by iteratively sampling new points and substituting those with the larger X (lower L) value, such that at each step the edge becomes the new λ threshold and the resampled live point satisfies the new $L > \lambda$ condition. Therefore, progressively the iso-likelihood contours are narrowed down (see Fig. 3.2) and the re-sampling algorithm is what differentiates the various implementations of nested sampling. For each iteration k, we can define the encoded prior mass at that iteration as X_k, which will correspond to the edge likelihood value L_k. If new samples are uniformly generated from progressively smaller nested regions, at each step the encoded prior mass shrinks on average by $e^{1/N}$ and assuming no uncertainties we could crudely assign $X_k = e^{-k/N}$. Alternatively, X_k can be sampled by setting $X_k = t_k X_{k-1}$, being t_k a random number within $(0, 1)$ drawn from $P(t_k) = N t_k^{N-1}$.

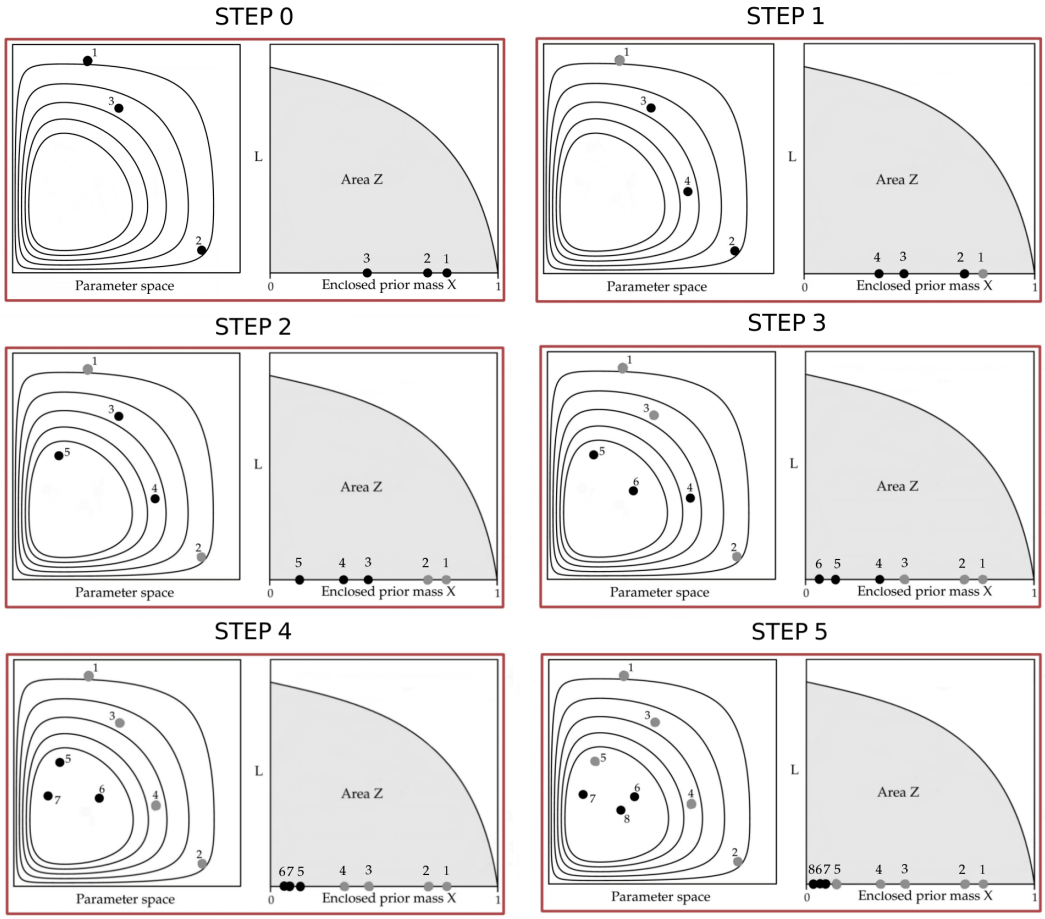

Figure 3.2: This sequence of panels shows five consecutive nested sampling steps. Each panel shows on the right the likelihood function as a function of X for a simplified model, the area below that function is the evidence $\mathcal{Z}$, while on the left it shows the sampled points in the parameter space. As we sample, the point with highest X becomes grey, and it acts as the new edge interval for sampling a new point. Five iterations with three live points (first panel, step 0) yields eight samples in total. Images taken and adapted from Skilling (2006).

By conveniently defining stripes of width $w_k = X_k - X_{k-1}$, the evidence integral can be obtained as

$$\mathcal{Z} = \sum_k w_k L_k \ . \tag{3.15}$$

Finally, at each X_k we can define the posterior as being $p_k = \frac{L_k}{\mathcal{Z}}$.

3.2 Extensions to the standard framework

3.2.1 Data compression

First attributed to Richard Bellman (Bellman, 1957; Bellman and Kalaba, 1959), the term *curse of dimensionality* refers to the fact that when the number of dimensions in a problem increases, the amount of data needed to maintain the same accuracy increases more rapidly, exponentially in most cases. This is often the case in cosmological analyses, for which dimensionality reduction is required prior to inference. In what follows, I provide examples of data compression algorithms I used, which are either based on analytical approximations or on the use of Machine Learning. In general, the data compression task can be recast as a problem of achieving a dimensionality which is as low as possible and still encodes the largest amount of information we are interested in, ideally preserving all of it. To this aim, it is important to keep in mind that some of these algorithms require the existence of a generative model, which must be good enough to ensure the relevant information is captured in the first place.

Relevant to this work there are in particular two different data compression algorithms I here briefly mention and will use in Chapter 4 and Chapter 6, respectively:

- *Score compression*. This analytical approach (Alsing and Wandelt, 2018) is based on the idea that, if we Taylor expand the log-likelihood $\mathcal{L}$ around the peak, at linear order the term that will contain the coupling between data and parameters will only be the score function $\boldsymbol{s} = \nabla \mathcal{L}_*$, namely the gradient of the log-likelihood evaluated at some fidu-

cial set of parameters $\boldsymbol{\theta}_*$. For N sampled parameters $\boldsymbol{\theta}$, the gradient will be of the form $\nabla = (\partial/\partial\theta_1, ..., \partial/\partial\theta_N)$ and the compressed vector will hence be given by as many components. This data compression scheme is applied to a science case in Chapter 6, where a more detailed description is given.

- *Regression networks*. Among all applications of ML, data compression algorithms are widely used in Astrophysics, from variational auto-encoders (Portillo et al., 2020) to neural networks (Charnock et al., 2018). Particularly relevant for the project presented in Chapter 4 are regression networks (Bishop, 2006). A simplistic scheme of a neural network is given in Fig. 3.3. The scheme can be divided into three main layers: the *input*, *hidden* and *output layers*. Generically speaking, the layer (l) (for $l \geq 1$, as $l = 0$ corresponds to the input

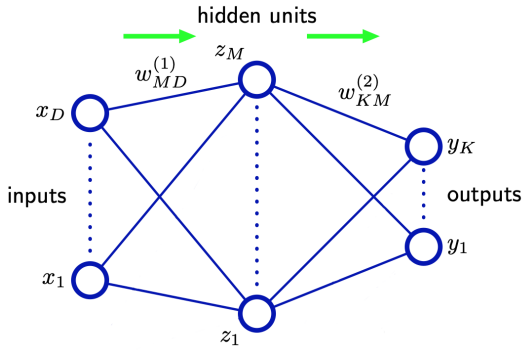

Figure 3.3: Schematic representation of a single hidden layer neural network, where $\boldsymbol{x}$ are the input features and $\boldsymbol{y}$ the output (prediction) of the network, while $\boldsymbol{z}$ generically indicate the intermediate variables in the hidden layer. To each of these variables is assigned a set of nodes. The nodes of each layer are combined via the weights $\boldsymbol{w}$ (see Eq. 3.16) to produce the nodes of the subsequent layer. Image taken from Bishop (2006).

layer) is composed of $M^{(l)}$ units[1] and to each of them is allocated a non-linear *activation function* $h^{(l)}(\cdot)$ of the inputs $z_i^{(l-1)}$ from the previous layer as

$$z_j^{(l)} = h^{(l)}\left(a_j^{(l)}\right) = h\left(\sum_{i=1}^{M^{(l-1)}} w_{ji}^{(l)} z_i^{(l-1)} + w_{j0}^{(l)}\right), \tag{3.16}$$

[1]Defined as *hidden units* if they are part of a hidden layer.

where $j = 1, ..., M^{(l)}$, $w_{ji}^{(l)}$ are the *weights* and $w_{j0}^{(l)}$ the *biases*. If we are at the first hidden layer ($l = 1$), then the $z_i^{(l-1)}$ vector will be made of the inputs to the neural network and $M^{(l-1)}$ is their dimensionality, and if we are at the last layer then $z_j^{(l)}$ are the outputs of the network and $M^{(l)}$ is their dimensionality. This type of structure is defined as a *feed-forward neural network* and we refer to all architecture and setup parameters as *hyperparameters*. As a *training dataset* $\{\boldsymbol{x}^t, \boldsymbol{y}^t\}$ is fed in, during the training, the weights and biases change to minimise a chosen error function of the network output with respect to the known true output $\boldsymbol{y}^t$, a typical choice is the mean squared error. This same error provides a metric to the performance of the network and it is used to judge the goodness of the training by evaluating its predictions over another set of data, defined as the *test set*. In the process of fine-tuning the hyperparameters, there is an extra testing dataset defined as the *validation set*. In the context of data compression, such a neural network structure can be used for regression purposes. Namely, based on the idea that the generative model for some data is known, it is possible to train the network feeding in simulations as inputs to the network and the generative parameters as target outputs. This means that, when fed with data, then a regression network will provide an estimate of the parameters of interest.

To conclude, a data compression method is whatever kind of mapping from the data space to a smaller set of compressed summary statistics. Independently of its nature, either an analytical compression scheme or a Machine Learning algorithm, the basic requirement is to allow for the information of interest to be preserved.

3.2.2 Simulation-based Inference

Our ability to constrain parameters in the Bayesian framework can be limited by a missing analytical form for the likelihood function. Often in cosmology, we assume the sampling distribution is Gaussian, but sometimes that might be either a bad approximation or unfeasible to numerically compute. In these cases, a convenient way to approximate the likelihood function, or the posterior distribution, is based on forward-modelling. Assuming there exists a trustworthy model that

can be used to generate simulations for different model and nuisance parameters, synthetic data can indeed be used to numerically evaluate some target distribution. This is simulation-based inference (SBI) (see Cranmer et al. 2020 for a review).

Two common approaches to SBI are Approximate Bayesian Computation (ABC) (Rubin, 1984; Beaumont et al., 2002; Sisson et al., 2018) and density-estimation simulation-based inference (Fan et al., 2013; Papamakarios et al., 2019; Lueckmann et al., 2018; Alsing et al., 2018). In the simple case of rejection ABC, the target probability is estimated by generating samples and keeping those that are within a certain threshold ϵ from the observed dataset (left panel in Fig. 3.4), the lower the ϵ the better the approximation. The downside of this method is that it can be extremely inefficient, not only all data realisations that are outside of the ϵ interval are wasted, but also this whole procedure must start over given new data. Instead, density-estimation simulation-based inference, which we will refer to as DELFI[2] is based on the idea of estimating the target density in the full joint space of data and parameters and then evaluating it at the observed realisation. In particular, this latter method uses neural networks to fit parametric forms to the target probability distribution. Given the estimators task, if the dimensionality of the joint space of data and parameters is too high, prior data compression is needed to facilitate the computations. In this case then the estimation would happen in the joint (summaries $\boldsymbol{t}$, parameters) space. The final posterior can be obtained by estimating either the joint density $p(\boldsymbol{\theta}, \boldsymbol{t})$, the conditional $p(\boldsymbol{\theta}|\boldsymbol{t})$ or $p(\boldsymbol{t}|\boldsymbol{\theta})$, as extensively explained by Alsing et al. (2019), who see the advantage of the latter against the others and propose an implementation, called PYDELFI, which we apply in Chapter 4.

NEURAL DENSITY ESTIMATORS

As already mentioned, the idea behind neural density estimation is to parameterise the target distribution, in our case $p(\boldsymbol{t}|\boldsymbol{\theta}; \mathbf{w})$, with neural networks via their weights $\mathbf{w}$, training them on the joint (summaries, parameters) space. The PYDELFI implementation uses two classes of neural

[2] DELFI stands for Density-Estimation Likelihood-Free Inference, as prior to SBI this class of method was referred to as LFI.

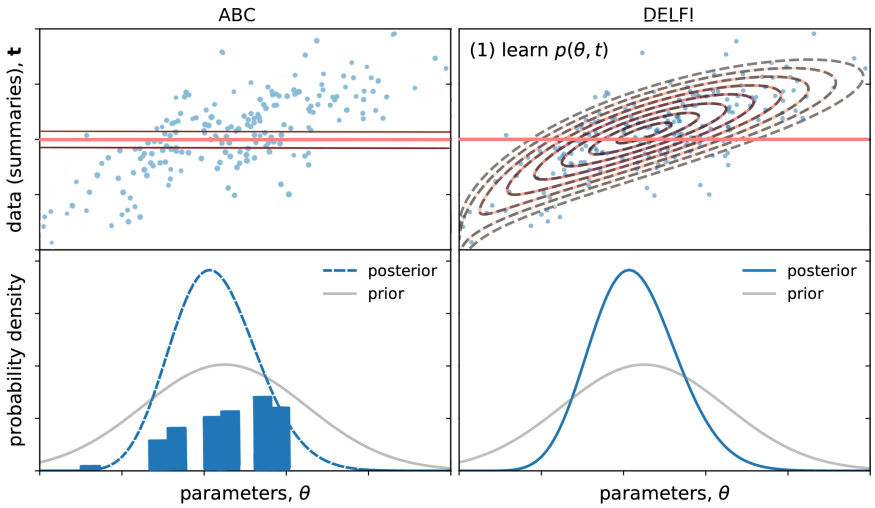

Figure 3.4: This figure is a visual comparison of the rejection ABC (left panel) and density estimation (right panel) algorithms. Given the (data, parameters) space, the ABC approach consists of generating simulations and accepting those that lie within ϵ (dark red lines in the upper part of the left panel) from the observation (orange line in the upper part of the panels). On the right panel instead, an example of density estimation is shown with target distribution being the joint density $p(\theta, t)$. Figure obtained modifying an example picture of Alsing et al. (2019).

density estimators (NDEs): mixture density networks (MDN; Bishop 1994) and masked autoregressive flows (MAFs; Papamakarios et al. 2019).

1) Mixture Density Networks These are mixture models, where the weights and properties of the K components are free functions of the parameters $\boldsymbol{\theta}$ and are parameterised in terms of the weights $\boldsymbol{w}$ of a chosen neural network. This means that if the model is a Gaussian mixture density model network, the likelihood function is modelled as (Alsing et al., 2019)

$$p(\boldsymbol{t}|\boldsymbol{\theta};\mathbf{w}) = \sum_{k=1}^{K} r_k(\boldsymbol{\theta};\mathbf{w})\mathcal{N}[\boldsymbol{t}|\boldsymbol{\mu}_k(\boldsymbol{\theta};\mathbf{w}), \boldsymbol{C}_k(\boldsymbol{\theta};\mathbf{w})] \ , \tag{3.17}$$

where each component has a relative weight $r_k(\boldsymbol{\theta};\mathbf{w})$, mean $\mu_k(\boldsymbol{\theta};\mathbf{w})$ and covariance $C_k(\boldsymbol{\theta};\mathbf{w})$. Referring to the schematic representation in Fig. 3.3, given a training set of $\{\boldsymbol{\theta}, \boldsymbol{t}\}$, the inputs are the parameters $\boldsymbol{\theta}$, while $r_k(\boldsymbol{\theta};\mathbf{w})$, $\mu_k(\boldsymbol{\theta};\mathbf{w})$ and $C_k(\boldsymbol{\theta};\mathbf{w})$ (for $k = 1, .., K$) are the outputs.

The network is trained to minimise $[-\sum_i \ln p(\boldsymbol{t}_i|\boldsymbol{\theta}_i;\mathbf{w})]$ and the structure is more complex than in the Figure. The larger the number of components the more flexible the model is.

II) MASKED AUTOREGRESSIVE FLOW A MAF is a stack of Masked Autoencoders for Density Estimation (MADEs; Germain et al. 2015), which are a particular example of neural autoregressive models (Uria et al., 2016). Specifically considering the target $p(\boldsymbol{t}|\boldsymbol{\theta})$, this can be factorised as

$$p(\boldsymbol{t}|\boldsymbol{\theta}) = \prod_{i}^{\dim(\boldsymbol{t})} p(t_i|t_{1:i-1},\boldsymbol{\theta}) \tag{3.18}$$

via the chain rule. The property of $p(t_i|t_{1:i-1},\boldsymbol{\theta})$ being dependent on the preceding random variables $t_{1:i-1}$, such that $p(t_i|t_{1:i-1},\boldsymbol{\theta})$, is called autoregressive property. Neural autoregressive models provide parametric fits to each $p(t_i|t_{1:i-1},\boldsymbol{\theta})$ via their weights, similarly to the previous case, such that $p(\boldsymbol{t}|\boldsymbol{\theta};\mathbf{w}) = \prod_i p(t_i|t_{1:i-1},\boldsymbol{\theta};\boldsymbol{w})$. Given the input parameters $\boldsymbol{\theta}$ and corresponding summaries $\boldsymbol{t}$, the MADE models each conditional as a Gaussian, so that all the means $\mu_i(\boldsymbol{\theta};\mathbf{w})$ and variances $\sigma_i(\boldsymbol{\theta};\mathbf{w})$ (for $i = 1,..,\dim(\boldsymbol{t})$) are functions of $(t_{1:i-1},\boldsymbol{\theta})$ parameterised by the weights $\boldsymbol{w}$, masked to satisfy the autoregressive property. For details refer to Germain et al. (2015).

The parameters of each conditional are combined to provide a unique output

$$u_i = \frac{t_i - \mu_i(t_{1:i-1},\boldsymbol{\theta};\boldsymbol{w})}{\sigma_i(t_{1:i-1},\boldsymbol{\theta};\boldsymbol{w})}\,, \tag{3.19}$$

which is the normalised value of the summary t_i. The idea behind MAFs is to construct a stack of MADES so that the $\boldsymbol{u}$ outputs of one MADE are the inputs to the next one, resulting in a larger flexibility; for any details refer to Papamakarios et al. (2019). As for MDNs, also this model is trained in order to minimise $[-\sum_i \ln p(\boldsymbol{t}_i|\boldsymbol{\theta}_i;\mathbf{w})]$.

In Chapter 4 we will test the reliability of the simulation-based inference algorithm in estimating the Hubble constant from some gravitational waves toy model. Given the latter, we run both the traditional analytical and the likelihood-free approaches, testing the second against biases. Instead, from Chapter 5 we will switch to Lyα forest science. We will test the cosmological information contained in the full correlation functions, beyond the BAO peak, by extending the traditional framework to allow for direct cosmological inference. Finally, in Chapter 6 we apply score compression on realistic Lyα correlation functions and explore some of its outcomes, including reliable covariance matrix estimation.

Part II

Simulation-based inference and data compression applied to cosmology

4 Unbiased likelihood-free inference of H_0 from light standard sirens

Multi-messenger observations of binary neutron star mergers offer a promising path towards resolution of the Hubble constant (H_0) tension, provided their constraints are shown to be free from systematics such as the Malmquist bias. In the traditional Bayesian framework, accounting for selection effects in the likelihood requires calculation of the expected number (or fraction) of detections as a function of the parameters describing the population and cosmology; a potentially costly and/or inaccurate process. This calculation can, however, be bypassed completely by performing the inference in a framework in which the likelihood is never explicitly calculated, but instead fit using forward simulations of the data, which naturally include the selection. This is Likelihood-Free Inference (LFI). Here, we use density-estimation LFI, coupled to neural-network-based data compression, to infer H_0 from mock catalogues of binary neutron star mergers, given noisy redshift, distance and peculiar velocity estimates for each object. We demonstrate that LFI yields statistically unbiased estimates of H_0 in the presence of selection effects, with precision matching that of sampling the full Bayesian hierarchical model. Marginalizing over the bias increases the H_0 uncertainty by only 6% for training sets consisting of $O(10^4)$ populations. The resulting LFI framework is applicable to population-level inference problems with selection effects across astrophysics.

The work presented in this chapter has been published, see Gerardi et al. (2021).

4.1 Introduction

In recent years, late-time measurements (Riess et al., 2021; Birrer et al., 2019; Wong et al., 2019) of the Hubble Constant, H_0, have diverged from estimates provided by early-time probes (Planck

Collaboration et al., 2020; Addison et al., 2018; Dark Energy Survey Collaboration and South Pole Telescope Collaboration, 2018; Philcox et al., 2020) (see Bernal et al. 2016; Verde et al. 2019; Bernal et al. 2021 for a summary). At the heart of the discrepancy is a 4.2σ tension between the latest direct measurement of $H_0 = (73.2 \pm 1.3)\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ by the SH0ES Team's Cepheid-supernova distance ladder (Riess et al., 2021) and the model-dependent value of $H_0 = (67.4 \pm 0.5)\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ inferred from observations of the cosmic microwave background (CMB) anisotropies by the Planck satellite (Planck Collaboration et al., 2020). While unforeseen systematic effects (Rigault et al., 2015; Jones et al., 2015; Rigault et al., 2020; Jones et al., 2018; Freedman et al., 2020; Brout and Scolnic, 2021) might be the cause of this disagreement, it is possible that this is a hint for new physics beyond the standard ΛCDM model (see Di Valentino et al. 2021 for a comprehensive summary of potential theoretical solutions). Despite considerable effort, however, no consensus on an explanation has been reached. This strongly motivates the need for a new, independent, direct probe of H_0. Gravitational waves (GWs) emitted by compact-object mergers — so-called *standard sirens* — are very promising in this regard (Schutz, 1986; Holz and Hughes, 2005; Dalal et al., 2006; Nissanke et al., 2010; Taylor et al., 2012; Messenger and Read, 2012; Nissanke et al., 2013; Oguri, 2016; Del Pozzo et al., 2017; Vitale and Chen, 2018; Seto and Kyutoku, 2018; Feeney et al., 2019; Gray et al., 2020; Feeney et al., 2021; Vitale et al., 2021), since their amplitude provides a self-calibrated estimate of the luminosity distance, d, depending only on General Relativity.

There are three types of compact-object systems typically considered for H_0 studies (Abbott et al., 2020): binary black holes (BBH), binary neutron stars (BNS) and neutron star - black hole (NSBH) systems. The potential for BNS and NSBH systems to have electromagnetic (EM) counterparts makes them particularly promising, as if an EM counterpart can be detected, the merger's host galaxy can be identified and its redshift measured, yielding H_0 when combined with d (Dalal et al., 2006; Nissanke et al., 2010; Nissanke et al., 2013; Vitale and Chen, 2018; Chen et al., 2018; Feeney et al., 2019; Seto and Kyutoku, 2018; Abbott et al., 2017). The first BNS system detected by the LIGO-Virgo Consortium, GW170817 (Abbott et al., 2017b), also produced an EM counter-

part (Abbott et al., 2017a), constraining H_0 to $70.0^{+12.0}_{-8.0}$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Abbott et al., 2017). The 10% constraints produced by this single event are expected to shrink to $\sim 1\%$ in the next 5-10 years once $O(100)$ events have been observed (Chen et al., 2018; Feeney et al., 2019; Seto and Kyutoku, 2018).

For standard siren estimates of H_0 to resolve the current tension, they must be shown to be free from systematic errors. Standard siren datasets suffer from Malmquist bias (Malmquist, 1922, 1925) which, left untreated, results in H_0 being overestimated. Traditional Bayesian methods must therefore take this effect into account by including in the likelihood terms involving the number (or, equivalently, fraction) of mergers that are expected to be detected given a set of population and cosmological parameters, $\bar{N}(\mathbf{\Omega})$ (Loredo, 2004; Abbott et al., 2017; Mandel et al., 2019; Mortlock et al., 2019; Vitale et al., 2021). The simplest method for calculating the expected number of detections is through Monte Carlo integration, i.e., repeated simulations of the dataset. Implementing this directly within a posterior sampling algorithm is, however, completely unfeasible, given the sheer number of simulations that would be needed. Instead, a single large catalogue of detected mergers can be generated using a fiducial set of population parameters and then reweighted to approximate $\bar{N}$ for any value of population parameters sampled (Tiwari, 2018). If the distribution of object parameters changes rapidly as a function of population parameters, however, a large (potentially computationally unfeasible) number of fiducial-population simulations are required to guarantee there are enough non-zero weights for the estimate of $\bar{N}$ to be reliable (the *effective* number of detected mergers must be at least four times the measured number (Farr, 2019)). Alternatively, $\bar{N}$ can be evaluated on a grid of $\mathbf{\Omega}$ and interpolated to generic population parameters (Mortlock et al., 2019; Feeney et al., 2021). While no reweighting is necessary in this case, the dependence on gridded computations means this method scales very poorly with parameter-set dimensionality.

Recently, Talbot and Thrane (2020) proposed a machine-learning based approach to this problem. The authors use a Gaussian mixture model to fit the distribution of object parameters found using a set of detected mergers drawn from a fiducial population. By dividing out the prior on

the object parameters for the fiducial population, they obtain an estimate of the probability of detecting a merger given its parameters. This estimate can be combined with the prior on the object parameters for a generic population to calculate $\bar{N}$ at any point sampled, either directly or via a neural-network-based interpolation. This approach suffers less bias than the reweighting method due to the assumption of a fiducial population, and comes at a cost of only $O(1000)$ simulated populations. However, the estimate of the detection probability as a function of object parameters is only defined over the range of parameters supported by the fiducial population; should this range change rapidly with the population parameters, the method's $\bar{N}$ estimates will lose accuracy.

Here, we take a different approach, demonstrating that the computation of $\bar{N}$ can be completely bypassed using Likelihood-Free Inference (LFI), which requires no analytic knowledge of the likelihood function. Specifically, we use Density-Estimation LFI (DELFI) (Papamakarios et al., 2019; Lueckmann et al., 2018; Alsing et al., 2018, 2019), in which the distribution of data as a function of the parameters that generated them is fit by supplying density estimators with a training set of simulated datasets. This fit is then used as a proxy likelihood to obtain posteriors on the parameters of interest. As the simulated data include the selection function, LFI automatically accounts for the Malmquist bias.

LFI's ability to accelerate the inference of the properties of individual BBH mergers has been demonstrated in a number of recent works (George and Huerta, 2018; Shen et al., 2021; Gabbard et al., 2021; Chua and Vallisneri, 2020; Green et al., 2020; Green and Gair, 2021; Delaunoy et al., 2020). Here, we apply LFI to population-level inference, taking as our example the inference of H_0 from 100 simulated GW-selected BNS mergers with EM counterparts. In this particular setting, traditional Bayesian inference (with $\bar{N}$ interpolated from a grid of cosmological values (Mortlock et al., 2019)) is feasible, and we take this approach as a ground truth from which we can robustly quantify any systematic errors introduced by LFI. We take as our inputs sets of individual mergers' observed redshifts, distances (generated via traditional (Abbott et al., 2017b) or likelihood-free analyses (e.g. Green and Gair, 2021)) and peculiar velocities, performing our

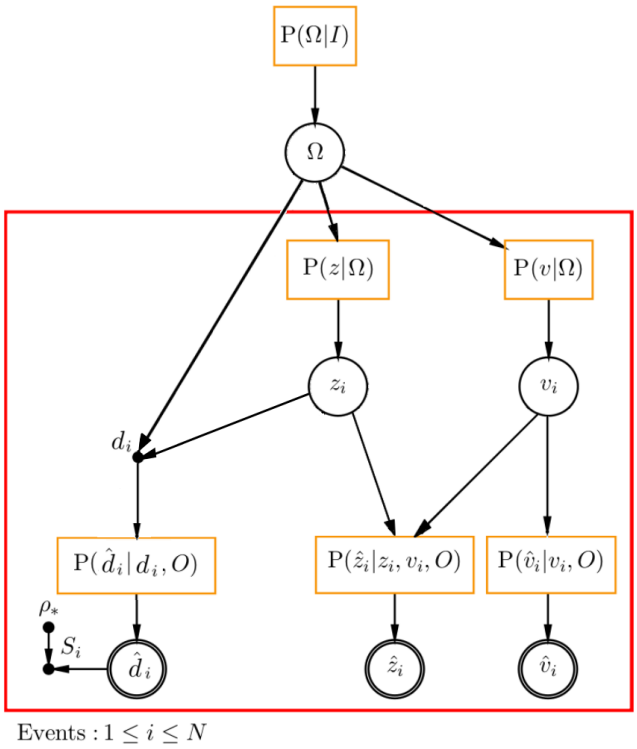

Figure 4.1: The hierarchical model used to describe our BNS population and data, adapted from Mortlock et al. (2019). Read top-to-bottom, parameters (circles) are drawn from probability distributions (orange rectangles) to generate observed quantities (double circles). I represents the prior information assumed about the cosmological parameters, $\mathbf{\Omega} = [H_0, q_0]$, and quantities within the red plate are specific to an individual merger.

LFI analysis with the aid of Pydelfi (Alsing et al., 2019). While we concentrate here on the inference of H_0 from BNS, the technique is applicable to population studies in general (e.g. The LIGO Scientific Collaboration et al., 2021; Kim, 2021).

We describe the hierarchical model we use to simulate our BNS mergers in Sect. 4.2, and explain our inference method in Sect. 4.3, highlighting the importance of data compression. Results are discussed in Sect. 4.4, and conclusions are drawn in Sect. 4.5.

4.2 Simulations

In this work we assume we possess noisy estimates of redshift $\hat{z}$, distance $\hat{d}$ and peculiar velocity $\hat{v}$ for each BNS merger. The mergers' $[\hat{\boldsymbol{z}}, \hat{\boldsymbol{d}}, \hat{\boldsymbol{v}}]$ are generated via the hierarchical model in Fig. 4.1,

which is loosely based on the model used in Mortlock et al. (2019). We assume that the strain data have been pre-compressed into estimates of $\hat{\boldsymbol{d}}$, which can be done rapidly using the likelihood-free method of Green and Gair (2021). Given the aforementioned prospects for solving the H_0 tension, we fix the number of mergers to $N = 100$. We consider two test cases, both assuming the same set of observables, but distinguished by whether GW selection is applied. Considering these two cases allows us to differentiate the impact of LFI alone from LFI specifically in the presence of selection effects.

In the following we wish to infer two cosmological parameters — the Hubble constant, H_0, and the deceleration parameter, q_0 — which we denote by $\boldsymbol{\Omega} = [H_0, q_0]$. For a given choice of $\boldsymbol{\Omega}$, true redshifts are randomly sampled from

$$\begin{aligned} P(z_i|\boldsymbol{\Omega}, z_{\max}) &= \frac{1}{(1+z_i)}\frac{dV}{dz}(\boldsymbol{\Omega})\mathcal{H}(z_{\max} - z_i) \\ &\simeq \frac{4\pi}{(1+z_i)}\frac{c^3 z^2}{H_0^3}[1 - 2(1+q_0)z_i]\mathcal{H}(z_{\max} - z_i), \end{aligned} \tag{4.1}$$

where $\mathcal{H}$ is a Heaviside step function. The final line is a good approximation for $z_{max} \ll 1$. Given a single cosmological redshift draw, the i^{th} distance is given by (Visser, 2004)

$$d_i(z_i, H_0, q_0) = \frac{cz_i}{H_0}\left[1 + \frac{1}{2}(1 - q_0)z_i\right]. \tag{4.2}$$

Denoting as $\mathcal{N}(\mu, \sigma)$ the normal distribution of mean μ and standard deviation σ, peculiar velocities are sampled from

$$\begin{aligned} P(v_i) &= \mathcal{N}(\mu_{v_\parallel}, \sigma_{v_\parallel}) \\ &= \mathcal{N}(0 \text{ km s}^{-1}, 500 \text{ km s}^{-1}). \end{aligned} \tag{4.3}$$

We convert our true redshifts, distances and peculiar velocities into observed quantities $\hat{\boldsymbol{x}} = [\hat{\boldsymbol{z}}, \hat{\boldsymbol{d}}, \hat{\boldsymbol{v}}]$ assuming, for simplicity, the marginal likelihoods are Gaussian[1], as follows

$$P(\hat{z}|z,v) = \mathcal{N}(z + v/c, \sigma_{\hat{z}} = 1.2 \times 10^{-3}) \tag{4.4}$$

$$P(\hat{d}|d) = \mathcal{N}(d, \sigma_{\hat{d}} = d/10) \tag{4.5}$$

$$P(\hat{v}|v) = \mathcal{N}(v, \sigma_{\hat{v}} = 200 \text{ km s}^{-1}). \tag{4.6}$$

When GW selection is not applied, we simulate populations by simply drawing from the above distributions N times. When using GW selection, we require that the signal-to-noise ratio (SNR), defined as

$$\rho_i(\hat{d}_i) = 12 \left(\frac{250\,\text{Mpc}}{\hat{d}_i} \right), \tag{4.7}$$

is greater than $\rho_* = 12$ for $i = [1, N]$. Introducing the GW selection changes the distribution of GW sources, reducing the effective upper redshift limit in a cosmology-dependent way, as shown in Fig. 4.2; the peak of the redshift distribution broadens and shifts to higher z for increasing H_0, while q_0 has a much smaller impact over this redshift range. For values of $H_0 \in [60, 80]$ km s^{-1} Mpc^{-1} and $q_0 \in [-2, 1]$, the redshift distribution is peaked at $z \simeq 0.05$. To ensure we generate sources at similar redshifts for our selection and no-selection populations (and consequently obtain similar constraints on cosmological parameters) we set $z_{\max}$ equal to 0.05 and 0.13 for the no-selection and selection cases, respectively.

4.3 Method

4.3.1 Traditional Inference

We begin by outlining the traditional approach to inferring parameters from GW-selected populations, before describing our adopted likehood-free methodology. The traditional framework has been set out in numerous references (Schutz, 1986; Dalal et al., 2006; Nissanke et al., 2010;

[1] The $\sigma_{\hat{d}} \propto d$ scaling of the distance uncertainty is chosen for simplicity. A better motivated choice would be $\sigma_{\hat{d}} \propto d^2$, given that the signal-to-noise ratio, Eq. 4.7, scales as $1/d$ (e.g., Mortlock et al., 2019).

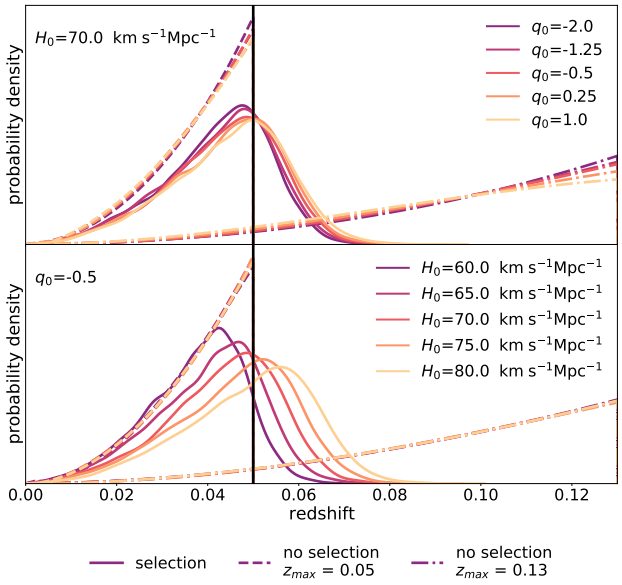

Figure 4.2: The dependence of BNS redshift distributions on q_0 (top) and H_0 (bottom) for our no-selection (dashed) and selection datasets (solid). To obtain comparable constraints on H_0 from the two datasets, we impose a cutoff at $z_{max} = 0.05$ for the no-selection case, while using $z_{max} = 0.13$ for the selection case. The input distribution for the selection dataset is shown as a dot-dashed line.

Taylor et al., 2012; Nissanke et al., 2013; Abbott et al., 2017; Mandel et al., 2019; Chen et al., 2018; Feeney et al., 2019; Gray et al., 2020; Vitale et al., 2021; Mortlock et al., 2019), but we will follow the notation of Mortlock et al. (2019) here. For simplicity, in this work we set aside the inference of the BNS properties (e.g. the NS mass distribution) and focus on the cosmology. As we are considering a fixed sample size here, the posterior on the cosmological parameters given a catalogue $\hat{\boldsymbol{x}} = [\hat{\boldsymbol{z}}, \hat{\boldsymbol{d}}, \hat{\boldsymbol{v}}]$ can be written as

$$P(\boldsymbol{z}, \boldsymbol{v}, H_0, q_0|\hat{\boldsymbol{x}}) \propto \frac{P(H_0)P(q_0)}{\left[\bar{N}(H_0, q_0)\right]^N} \times \prod_{i=1}^{N} P(z_i|H_0, q_0, z_{\rm max})P(v_i)P(\hat{z}_i|z_i, v_i)P(\hat{d}_i|d_i)P(\hat{v}_i|v_i). \tag{4.8}$$

We assume truncated Gaussian priors on the cosmological parameters

$$\begin{aligned} P(H_0) &= \mathcal{H}(H_0 - 60)\mathcal{H}(80 - H_0) \\ &\quad \times \mathcal{N}\left(70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}, 20\ \mathrm{km\,s^{-1}\,Mpc^{-1}}\right) \\ P(q_0) &= \mathcal{H}(q_0 + 2)\,\mathcal{H}(1 - q_0)\,\mathcal{N}(-0.55, 0.5). \end{aligned} \tag{4.9}$$

All other distributions are taken to match those set out in Sect. 4.2.

The impact of the selection function is captured by the factor of $\left[\bar{N}(H_0, q_0)\right]^{-N}$. $\bar{N}$ (which, recall, denotes the expected number of *detected* mergers) must be evaluated at every point in parameter space sampled by a particular inference tool. Here, we follow Mortlock et al. (2019) in evaluating $\bar{N}$ on a 10×10 grid in H_0 and q_0 (boosting the fiducial detection rate $\Gamma = 1540\mathrm{Gpc^{-3}yr^{-1}}$ (Abbott et al., 2017b) by a factor of 130 to reduce sample variance), and then fitting using a fourth-order (15-coefficient) polynomial. Following Mortlock et al. (2019), we then perform traditional Bayesian Inference using No-U-Turn-Sampling (Hoffman and Gelman, 2011) as implemented in the PyStan package (Carpenter et al., 2017; Stan Development Team, 2018), explicitly sampling each merger's true redshift and peculiar velocity along with H_0 and q_0. We take the marginal posteriors on H_0 and q_0 output by PyStan as the ground truth in the tests that follow.

4.3.2 Likelihood-Free Inference

Explicitly calculating $\bar{N}(H_0, q_0)$ at each point of parameter space sampled is computationally unfeasible. The methods proposed to circumvent this issue must balance computational cost and accuracy. The standard method of estimating $\bar{N}$ via a reweighted sum over a set of detected mergers generated using a fiducial population (Tiwari, 2018; Farr, 2019; The LIGO Scientific Collaboration et al., 2021) works well provided the object-level parameter distribution for generic population parameters does not differ too strongly from that of the fiducial population (Farr, 2019). To counter this, the fiducial detected merger population must be oversampled, increas-

ing the cost of both generating the detected sample and evaluating the likelihood. The cost of the former will become prohibitive in any setting where the distributions of object parameters have finite (or strongly suppressed) support which changes with the population parameters. Talbot and Thrane (2020) estimates $\bar{N}$ by fitting the distribution of object parameters found in the fiducial detection set and from this obtaining an estimate of the probability of detecting a merger given its parameters. This reduces both the computational cost and the bias due to estimating the detection probability from a fiducial population that might differ strongly from the underlying truth; however, it still fundamentally depends on the assumption of a fiducial population. The gridded approximation (Mortlock et al., 2019) we use for our traditional Bayesian analysis here does not require a fiducial population but is computationally expensive, requiring $\sim 130 \times N$ selected mergers for each single point of the grid, hence ~ 13000 detected samples in total. It can not be scaled to problems with a large number of population parameters.

Here we demonstrate that we can bypass the $\bar{N}$ calculation entirely using likelihood-free methods, which are based solely on simulations and therefore naturally account for selection effects. In particular, we use Density-Estimation Likelihood Free Inference (DELFI) (Papamakarios et al., 2019; Lueckmann et al., 2018; Alsing et al., 2018, 2019), in which synthetic mergers sampling the joint parameter-data space $(\boldsymbol{\Omega}, \hat{\boldsymbol{x}})$ are used to train neural density estimators (NDEs) to fit $P(\hat{\boldsymbol{x}}|\boldsymbol{\Omega})$, the probability of obtaining GW-selected data given the population parameters. By fitting this distribution, we implicitly marginalize over the mergers' true redshifts and peculiar velocities. The fit is evaluated at the observed data $\hat{\boldsymbol{x}}_{\rm obs}$ to obtain $P(\hat{\boldsymbol{x}}_{obs}|\boldsymbol{\Omega}; \mathbf{w})$, a parametric model for the likelihood depending on the trained weights $\mathbf{w}$ of the neural density estimators. This is then multiplied by the prior to yield the final posterior $P(\boldsymbol{\Omega}|\hat{\boldsymbol{x}}_{obs}) \propto P(\boldsymbol{\Omega})P(\hat{\boldsymbol{x}}_{obs}|\boldsymbol{\Omega}; \mathbf{w})$.

Our LFI analysis uses PYDELFI[2], an implementation of DELFI developed by Alsing et al. (2019), based on Papamakarios et al. (2019), Lueckmann et al. (2018) and Alsing et al. (2018). PYDELFI learns a parametric model to the conditional distribution $P(\hat{\boldsymbol{x}}|\boldsymbol{\Omega})$ — via *on-the-fly* or precomputed simulations — using a set of NDEs. The NDE components can be freely chosen as a com-

[2] https://github.com/justinalsing/pydelfi

bination of mixture density networks (MDNs) and masked autoregressive flows (MAFs) (see Alsing et al. 2019; Bishop 1994; Papamakarios et al. 2017; Papamakarios et al. 2019 for details on the NDEs). To reduce the possibility of pathological behavior from one particular NDE affecting our results, we create an ensemble of estimators by stacking together five MDNs (with one to five Gaussian components) and one MAF. We use the same ensemble of NDEs for all PYDELFI runs. To reduce variance in our results, we train all of the NDEs using a fixed set of 2000 simulated training populations, rather than letting the algorithm generate on-the-fly simulations. These training samples are obtained by uniformly drawing from $H_0 \in [60, 80]$ km s^{-1} Mpc^{-1} and $q_0 \in [-2, 1]$. The choice of the training-set size is empirically driven by the estimators' efficiency: there exists a (setting-specific) limiting training-set size beyond which there is no significant improvement in the training (Alsing et al., 2019). Reducing the training set to 1000 populations significantly impacts the quality of our results; boosting it to 10000 does not improve the results enough to justify the higher computational cost.

DATA COMPRESSION METHOD

As the simulated catalogues consist of $N = 100$ sources, performing LFI on the raw data would require fitting a 302-dimensional probability distribution, which is unfeasible (given the available resources in terms of number of simulations and our fidelity requirements). In order to reduce the dimensionality of the inference space, the data must be compressed to a set of summary statistics $\hat{\boldsymbol{t}}$. While there is no general requirement over the dimensionality of the compressed vector, we specifically require for our purposes the latter to be a vector of $\dim(\hat{\boldsymbol{t}}) \equiv \dim(\boldsymbol{\Omega})$ components (i.e., one compressed summary per parameter of interest). Identifying suitable summary statistics translates into finding a map $f : \hat{\boldsymbol{x}} \to \hat{\boldsymbol{t}}$ that compresses the data while retaining as much information as possible. Methods capable of performing such a mapping include score compression (Alsing and Wandelt, 2018; Alsing et al., 2018; Alsing and Wandelt, 2019a), Information Maximizing Neural Networks (Charnock et al., 2018) and regression neural networks (NNs) (Bishop, 2006). In this work, we train regression neural networks to compress generic merger data into

estimates of the generative cosmological parameters. For training purposes, we need to construct a set of training and validation datasets, for which the underlying cosmology is known and will constitute the target. The network will ultimately compress the noisy data to a set of summary statistics which correspond to a prediction about the generative cosmological model. To avoid any dependence on the particular training initialization of a single network, we create an ensemble of 9 trained neural networks, all defined by the same settings and trained on the same exact data but using different random initial weights.

The raw observables span a broad range of magnitudes — $\hat{z} \simeq O(10^{-2})$, $\hat{d} \simeq O(10^{2})$ and $\hat{v} \simeq O(10^{3})$ — which can cause problems in the training process. If there are large differences in scale between different components of the data vector, the NN will naturally prioritize the larger components, effectively ignoring part of the dataset. Moreover, the magnitude of the data vector determines the update rate, so large values might lead to stability problems. Prior to feeding data into any neural network, therefore, we normalize the data to ensure they are all at roughly the same scale. We first sort all merger catalogues by redshift to reduce the variability to which each NN input node is exposed. We then concatenate each catalogue's $\hat{\boldsymbol{z}}$, $\hat{\boldsymbol{d}}$ and $\hat{\boldsymbol{v}}$ to create a single 300-element raw-input vector. Finally we shift and scale by the mean and standard deviation of 100 catalogues generated at our fiducial cosmology $[H_0, q_0] = [70, -0.5]$ to create the normalized inputs for our regression networks. We also normalize the target parameters which generated the training and validation datasets, shifting and scaling their distributions to be within 0 and 1. The NN predictions — our summary statistics — are hence normalized estimates of the cosmological parameters.

Data compression optimization

The choice of architecture and settings for our neural networks is completely free, which poses an intimidating optimization problem over the vast number of possible NN architectures and settings. To define a NN we must choose an architecture, its activation function and training, by tuning batch size, learning rate and potentially employing regularization methods. We can-

not reasonably explore all of these choices, and we therefore consider neural networks composed of two hidden layers, each made of 128 hidden units, fix the activation function to be a Leaky ReLU (Maas et al., 2013) with `alpha` $= 0.01$,[3] and focus on finding the best combination of batch size n_{batch} and learning rate α from a small set of choices, namely $n_{batch} = [100, 500]$ and $\alpha = [10^{-4}, 5 \times 10^{-4}, 10^{-3}]$. To avoid potential overfitting, we consider regularization terms, which control the training while acting on the loss function, set to be the mean squared error (MSE). We toggle between Ridge and Lasso regression methods, which use L2 and L1 regularizations respectively (Hastie et al., 2009), and explore a few values of the parameters weighting the regularization term, $\lambda_{1,2}$, namely $\{\lambda_{1,2} = 0\}$,$\{\lambda_1 = 0, \lambda_2 = [10^{-4}, 2 \times 10^{-4}]\}$ and $\{\lambda_1 = [10^{-4}, 2 \times 10^{-4}], \lambda_2 = 0\}$. We define the optimal compressor as the NN for which PyDELFI most faithfully reproduces PyStan's results for a range of $[H_0, q_0]$. The process by which we determine the optimal NN settings is described in the following.

For each combination of batch size, learning rate and regularization, we first train the regression NN on a set of n_{train} samples of known cosmology, validating with a further n_{val} datasets. To determine the impact of the amount of training data available on the final inference, we consider two training set sizes, the first with $[n_{train}, n_{val}] = [5000, 2000]$ and the second with $[500000, 100000]$. In all cases, the generative cosmologies are sampled from $H_0 \in [60, 80]\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $q_0 \in [-2, 1]$ using the Latin hypercube method.

To determine the NN parameters that optimize LFI performance for a range of underlying cosmologies, we generate 100 test catalogues for cosmological parameters sampled from $H_0 \in [65, 75]\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $q_0 \in [-0.7, -0.3]$ using the Latin hypercube method (the reason for this restricted range will be explained in Sect. 4.4). We then perform traditional Bayesian inference and LFI on each test catalogue, for each choice of NN parameters. Given these results, we compute the differences $b_{H_0} = \hat{H}_0^{trad} - \hat{H}_0^{LFI}$ and $b_{q_0} = \hat{q}_0^{trad} - \hat{q}_0^{LFI}$ between the maximum-posterior estimates of the cosmological parameters from the traditional and LFI ap-

[3] https://keras.io/api/layers/activation_layers/leaky_relu/

proaches, which we define as "biases"[4]. Compiling the results from all of the test catalogues, we calculate the means ($\bar{b}_{H_0,q_0}$) and standard deviations ($\sigma_{b_{H_0,q_0}}$) of the biases injected by LFI for each compression NN. The optimal compression network is chosen to be that which minimizes the standard deviation of the H_0 bias, provided its mean bias is consistent with zero.

In addition to requiring LFI produces unbiased estimates of the cosmological parameters, we also want to ensure our compression is as lossless as possible, i.e., that the LFI and traditional constraints have similar H_0 uncertainties. To do so, we need the total uncertainty in the LFI parameter constraints, which we approximate as the quadrature sum of the "raw" uncertainty of the LFI posteriors and the additional uncertainty due to the bias.[5] We estimate the former by calculating the mean variance of the LFI cosmological parameter posteriors over all 100 test catalogues; the uncertainty on the bias is simply $\sigma_{b_{H_0}}$. Hence, the increase in the H_0 uncertainty expected from replacing traditional Bayesian inference with LFI in this setting can be estimated by calculating

$$\%\hat{\sigma}^{H_0}_{incr} = 100 \times \left(\frac{\sqrt{\left(\sigma^{H_0}_{\mathrm{LFI}}\right)^2 + \sigma^2_{b_{H_0}}}}{\sigma^{H_0}_{trad}} - 1 \right). \tag{4.10}$$

4.4 Results

We first consider the no-selection case to demonstrate the feasibility of LFI in this setting and obtain a baseline for its impact on the precision and accuracy of the inference. We then add in GW selection to determine whether selection specifically affects LFI's performance, and to provide a final estimate of the systematics.

[4] As such, these biases contain contributions from any inaccuracies in the traditional $\bar{N}$ estimation and inference (expected to be small) and loss of information through imperfect compression. If the compression is lossless and the PyStan inference introduces no error, the PyStan and LFI posteriors should match perfectly.

[5] This is equivalent to marginalizing over an unknown additive bias, assuming the parameters and bias are independent and Gaussian-distributed.

n_{batch}	α	regularizer	b_{H_0} [km s^{-1} Mpc^{-1}]	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
NO SELECTION CASE							
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [5000, 2000]$							
100	10^{-4}	$\lambda_1 = 10^{-4}$	0.024 ± 0.35	-0.003 ± 0.095	1.014 ± 0.045	0.95 ± 0.035	7.64%
	5×10^{-4}	$\lambda_1 = 10^{-4}$	-0.002 ± 0.365	0.004 ± 0.098	1.028 ± 0.042	0.952 ± 0.032	9.43%
	10^{-3}	$\lambda_1 = 10^{-4}$	0.007 ± 0.352	0.009 ± 0.09	1.024 ± 0.048	0.952 ± 0.038	8.58%
500	10^{-4}	$\lambda_1 = 10^{-4}$	0.012 ± 0.358	-0.003 ± 0.092	1.003 ± 0.043	0.947 ± 0.036	6.81%
	5×10^{-4}	$\lambda_1 = 10^{-4}$	0.026 ± 0.328	0.001 ± 0.091	1.018 ± 0.051	0.948 ± 0.026	7.3%
	10^{-3}	$\lambda_1 = 10^{-4}$	0.021 ± 0.322	-0.0 ± 0.087	1.012 ± 0.054	0.943 ± 0.036	6.45%
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [500000, 100000]$							
100	10^{-4}	$\lambda_2 = 2 \times 10^{-4}$	-0.073 ± 0.193	0.015 ± 0.061	0.979 ± 0.042	0.945 ± 0.038	-0.05%
	5×10^{-4}	–	-0.061 ± 0.218	0.014 ± 0.071	0.978 ± 0.048	0.948 ± 0.04	0.35%
	10^{-3}	–	-0.058 ± 0.21	0.02 ± 0.058	0.973 ± 0.042	0.945 ± 0.04	-0.35%
500	10^{-4}	$\lambda_2 = 10^{-4}$	-0.043 ± 0.193	0.017 ± 0.066	0.972 ± 0.041	0.944 ± 0.039	-0.77%
	5×10^{-4}	$\lambda_2 = 2 \times 10^{-4}$	-0.053 ± 0.208	0.015 ± 0.061	0.975 ± 0.046	0.944 ± 0.037	-0.15%
	10^{-3}	$\lambda_2 = 10^{-4}$	-0.062 ± 0.189	0.012 ± 0.06	0.979 ± 0.048	0.946 ± 0.035	-0.22%

Table 4.1: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for the NNs whose regularization choice minimizes the bias for each combination of batchsize n_{batch} and learning rate α in the no-selection case.

4.4.1 No-Selection Case

Considering the no-selection case first gives us a baseline for gauging LFI's performance in the more complex setting with selection, allowing us to determine whether selection specifically has any impact on LFI. We train our compression NNs for all combinations of the aforementioned batchsize, learning rate and regularizer choices, for both training-set sizes $[n_{train}, n_{val}]$. Each of these neural networks provides different compression performance and thus all are tested as compressors in the LFI workflow. An example of compression performance for $[n_{train}, n_{val}] = [500000, 200000]$ is given in Fig. 4.3, which shows the summary statistics $\hat{\boldsymbol{t}}$ output by the regression NN against the generative cosmological parameters for the validation set. Focusing on the $\hat{t}_1 - H_0$ and $\hat{t}_2 - H_0$ plots for now, we notice that the width and slope of the distribution change at the edges of the training set, shaded in grey. As the NN behaviour might be suboptimal in these ranges, we generate the test samples used to optimize the compressor settings from values of H_0 within $[65, 75]$ km s^{-1} Mpc^{-1}, lying in the unshaded area.

We identify the best regularization for each combination of batchsize and learning rate using the b_{H_0} distribution. The b_{H_0} and b_{q_0} probability densities are respectively shown as blue and orange violin plots in Fig. 4.4, for $[n_{train}, n_{val}] = [5000, 2000]$, and summarized in Table 4.1.

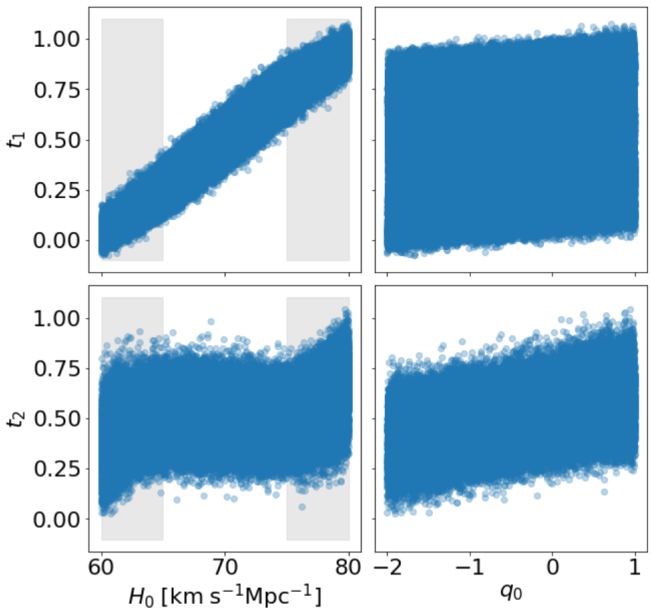

Figure 4.3: The summary statistics $\hat{\boldsymbol{t}} = (\hat{t}_1, \hat{t}_2)$ output by our compression NN plotted against the cosmological parameters at which the corresponding data were generated. This NN was trained with $[n_{\mathrm{batch}}, \alpha, \lambda_{1,2}] = [100, 10^{-4}, 0]$, and the points correspond to the validation dataset for the $[n_{train}, n_{\mathrm{val}}] = [500000, 200000]$ setup. The shaded areas indicate the regions of H_0 where the slopes of the summary statistics change with respect to the central trend.

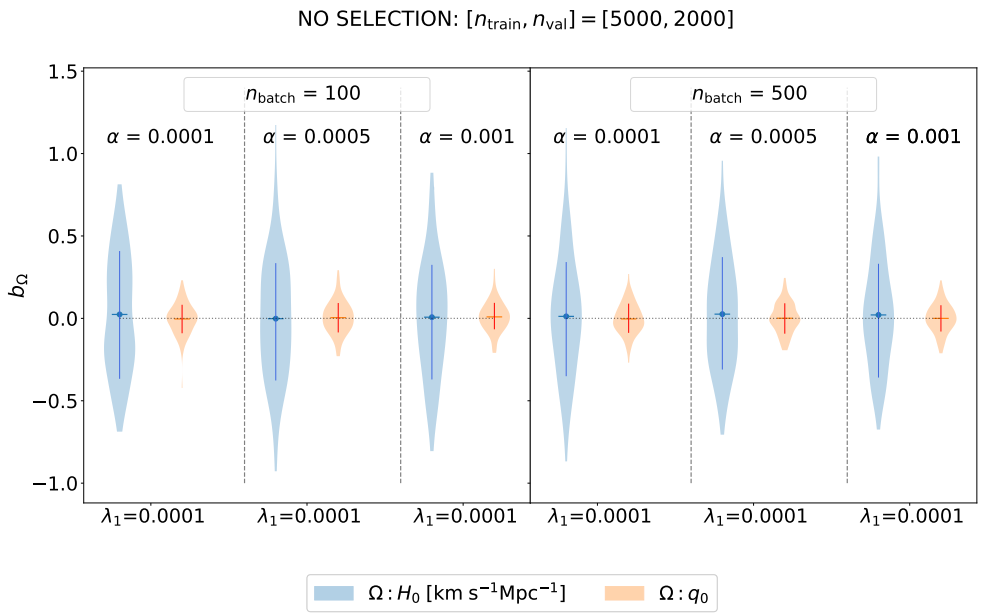

Figure 4.4: Violin plots for the $b_{H_0} = \hat{H}_0^{trad} - \hat{H}_0^{\text{LFI}}$ (blue) and $b_{q_0} = \hat{q}_0^{trad} - \hat{q}_0^{\text{LFI}}$ (orange) bias distributions for the no-selection setting. Results are shown for the NNs whose regularization choice minimizes the bias for each combination of batchsize n_{batch} and learning rate α. Dots represent the mean biases, and lines the 1σ errorbars. The mean biases are consistent with zero, and the bias distributions are considerably narrower than the relevant parameter posteriors, for all NNs plotted.

Results for all NN parameter choices can be found in Tables 4.6.1 and 4.6.2. From the violin plots we see that the likelihood-free inference of both H_0 and q_0 is unbiased, since the bias is consistent with zero for all choices of NN parameters. For the best models, independent of the specific NN parameters and data realization, LFI's maximum posterior estimate for both parameters is typically well within PyStan's 1σ posterior uncertainty ($\geq 0.89\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for these test populations).

We observe that for our smaller training set, regularization greatly improves performance. As an example, considering $[n_{batch}, \alpha] = [100, 10^{-4}]$ we find that adding a regularization term $\lambda_1 = 10^{-4}$ reduces $\sigma_{b_{H_0}}$ from 1.75 to 0.35 and markedly increases the H_0 constraining power, reducing $f_\sigma^{H_0} = \sigma_{H_0}^{LFI}/\sigma_{H_0}^{trad}$ from 1.95 to 1.06. With regularization added, the width of the LFI H_0 posterior is compatible with PyStan's. Considering the larger training set reduces the impact of the regularizer and significantly reduces the H_0 LFI posterior's uncertainty, which we find to be systematically $\sim 2-3\%$ smaller than PyStan's: we suspect that this is due to slight overfitting by Pydelfi. The LFI q_0 constraints are also $\sim 5\%$ tighter than PyStan's, independent of the size of the training set.

For the $[n_{train}, n_{val}] = [5000, 2000]$ setup, the network with $[n_{batch}, \alpha, \lambda_1] = [500, 10^{-3}, 10^{-4}]$ imparts the smallest bias in the H_0 posterior, with $\sigma_{b_{H_0}} = 0.32$. The H_0 bias shrinks further when using our larger training set, with $\sigma_{b_{H_0}} = 0.19$. As the bias is small and consistent with zero it could be ignored when doing population-level inference; here, however, we marginalize over it and find that it would impart a 6.45% and -0.05% increase in the quoted H_0 uncertainty, respectively: well within any reasonable tolerance. We note here that this slight increase in uncertainty is entirely down to imperfect compression, since in tests Pydelfi provides the same posteriors when rerunning on the same compressed data.

One advantage of using a regression neural network for compression is that it only relies on a fiducial model for the computation of the mean and standard deviations used to normalize the neural network inputs. Nevertheless, the compression is sensitive to the choice of the training and validation data, as well as the range of sampled $\mathbf{\Omega}$ values. To investigate the randomness of

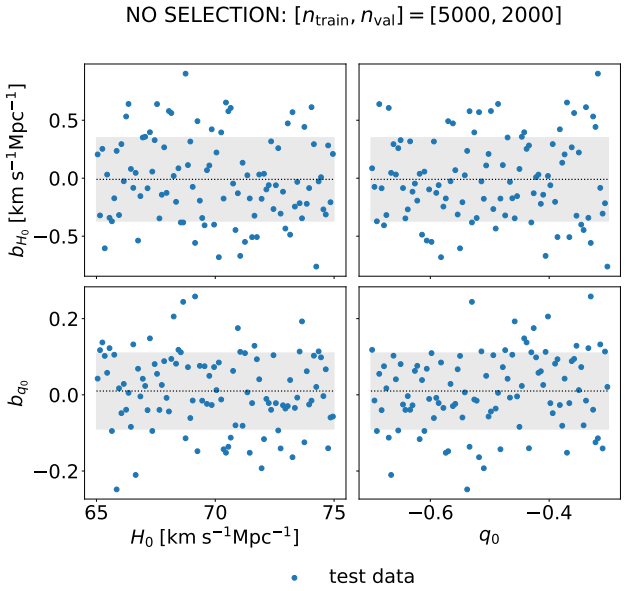

Figure 4.5: Distribution of generative parameters and LFI posterior biases. The one-sigma range of the bias is shaded grey. The neural network model used to perform the compression and generate this plot corresponds to the NN parameters combination $[n_{batch}, \alpha, \lambda_1] = [500, 10^{-3}, 10^{-4}]$ for $[n_{train}, n_{val}] = [5000, 2000]$.

the H_0 bias with respect to the sampled parameter space, we plot the biases against the generative parameters for all 100 test catalogues for our best compression network in Fig. 4.5. We find there is no major correlation between the true parameters and the biases (for example, for the best model of the $[n_{train}, n_{val}] = [5000, 2000]$ setup, we find correlation coefficients of $C(H_0, b_{H_0}) = -0.13$ and $C(q_0, b_{H_0}) = -0.023$).

4.4.2 Selection Case

We now proceed to determine the impact of selection on the compression. As in the no-selection case, we first optimize the regularization for each combination of batchsize and learning rate. We compute the distributions of the H_0 and q_0 biases, plotting the results for the best compressors in Fig. 4.6 and tabulating their performance in Table 4.2. Results for all the NN parameters can be found in Tables 4.6.3 and 4.6.4. As in the no-selection case, the LFI maximum-posterior parameter estimates are unbiased when compared to the PyStan baseline.

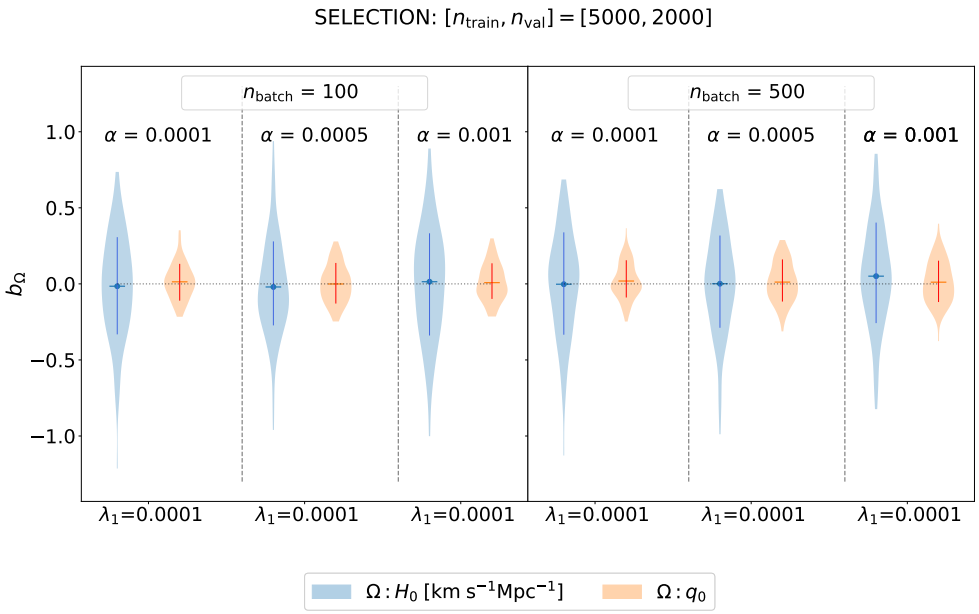

Figure 4.6: Violin plots for the b_{H_0} (blue) and b_{q_0} (orange) bias distributions for the setting with GW selection. Results are shown for the NNs whose regularization choice minimizes the bias for each combination of batchsize n_{batch} and learning rate α. Dots represent the mean biases, and lines the 1σ errorbars. As in the no-selection case, the mean biases are all consistent with zero, and the bias distributions are all considerably narrower than the relevant parameter posteriors.

SELECTION CASE							
n_{batch}	α	regularizer	b_{H_0} [km s^{-1} Mpc^{-1}]	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [5000, 2000]$							
100	10^{-4}	$\lambda_1 = 10^{-4}$	-0.015 ± 0.338	0.014 ± 0.115	1.013 ± 0.039	1.005 ± 0.044	6.53%
	5×10^{-4}	$\lambda_1 = 10^{-4}$	-0.02 ± 0.313	-0.001 ± 0.122	1.014 ± 0.041	1.005 ± 0.058	5.9%
	10^{-3}	$\lambda_1 = 10^{-4}$	0.014 ± 0.357	0.008 ± 0.119	1.018 ± 0.042	1.006 ± 0.051	7.67%
500	10^{-4}	$\lambda_1 = 10^{-4}$	-0.002 ± 0.334	0.019 ± 0.116	1.025 ± 0.04	1.01 ± 0.049	7.63%
	5×10^{-4}	$\lambda_1 = 10^{-4}$	0.001 ± 0.313	0.012 ± 0.128	1.025 ± 0.038	1.013 ± 0.036	6.99%
	10^{-3}	$\lambda_1 = 10^{-4}$	0.051 ± 0.329	0.011 ± 0.137	1.019 ± 0.053	1.011 ± 0.058	6.91%
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [500000, 100000]$							
100	10^{-4}	$\lambda_2 = 10^{-4}$	-0.032 ± 0.184	0.022 ± 0.092	0.976 ± 0.031	1.006 ± 0.036	-0.73%
	5×10^{-4}	$\lambda_2 = 10^{-4}$	-0.033 ± 0.177	0.02 ± 0.092	0.979 ± 0.039	1.003 ± 0.043	-0.56%
	10^{-3}	–	0.0 ± 0.183	0.026 ± 0.091	0.965 ± 0.03	1.004 ± 0.038	-1.88%
500	10^{-4}	$\lambda_1 = 10^{-4}$	-0.022 ± 0.178	0.015 ± 0.093	0.978 ± 0.036	1.003 ± 0.039	-0.7%
	5×10^{-4}	$\lambda_2 = 10^{-4}$	-0.013 ± 0.18	0.019 ± 0.086	0.977 ± 0.035	1.006 ± 0.038	-0.68%
	10^{-3}	$\lambda_2 = 2 \times 10^{-4}$	-0.01 ± 0.199	0.021 ± 0.083	0.979 ± 0.043	1.003 ± 0.042	-0.18%

Table 4.2: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for the NNs whose regularization choice minimizes the bias for each combination of batchsize n_{batch} and learning rate α in the selection case.

As before, for our smaller training set regularization overall largely improves the performance. Considering $[n_{batch}, \alpha] = [100, 10^{-4}]$ as an example as before, we find that regularizing the training for $\lambda_1 = 10^{-4}$ reduces the uncertainty on the H_0 bias from 1.71 to 0.34 and greatly improves the H_0 constraining power, from $f_\sigma^{H_0} = 1.77$ to 1.06. As in the no-selection case, the LFI posteriors produced using the optimal compressors are completely compatible with PyStan's. Again, increasing the training set size reduces the impact of the regularizer and significantly reduces the LFI H_0 posterior's uncertainty, to $\sim 2.5\%$ smaller than PyStan's.

For the $[n_{train}, n_{\rm val}] = [5000, 2000]$ setup two NN compressors minimize the H_0 bias, with $\sigma_{b_{H_0}} = 0.31$. These are defined by $[n_{\rm batch}, \alpha, \lambda_1] = \{[100, 5 \times 10^{-4}, 10^{-4}], [500, 5 \times 10^{-4}, 10^{-4}]\}$. As in the no-selection case the best models compressors use λ_1 regularization. For the larger $[n_{train}, n_{\rm val}] = [500000, 100000]$ setup, the smallest standard deviation for the H_0 bias is again considerably smaller: $\sigma_{b_{H_0}} = 0.18$ for the compressor with $[n_{batch}, \alpha, \lambda_2] = [100, 5 \times 10^{-4}, 10^{-4}]$. As in the no-selection case, we compute the percentage increase in uncertainty on H_0 imparted by replacing traditional inference with LFI, marginalizing over the bias. For the aforementioned three best compressors, these percentage increases are $\{5.9\%, 6.99\%\}$ and -0.56%, respectively, compatible with that determined for the no-selection case. Including GW selection does not impact LFI performance on a statistical level. Illustrative examples of the H_0-q_0 joint posteriors produced by Pydelfi and PyStan can be found in Fig. 4.6.1.

In Fig. 4.7 we plot the values of the b_{H_0} and b_{q_0} distributions against true input cosmology parameters. Unlike in Fig. 4.5, there is a clear dependence of b_{H_0} and b_{q_0} on the true value of H_0 that generated the data. The strongest correlation is between the q_0 bias and the generative H_0, with a correlation coefficient of -0.47 for the best model $[n_{\rm batch}, \alpha, \lambda_1] = [100, 5 \times 10^{-4}, 10^{-4}]$ of the smaller training set $[n_{train}, n_{\rm val}] = [5000, 2000]$. Increasing the size of the training sample generates stronger correlation (-0.66 for the best model). This indicates the regression (or, indeed, an imperfect fit to PyStan's $\bar{N}$) is not capturing the selection function perfectly, and that other compression methods may fare better. Nevertheless, for the optimal compressors the biases

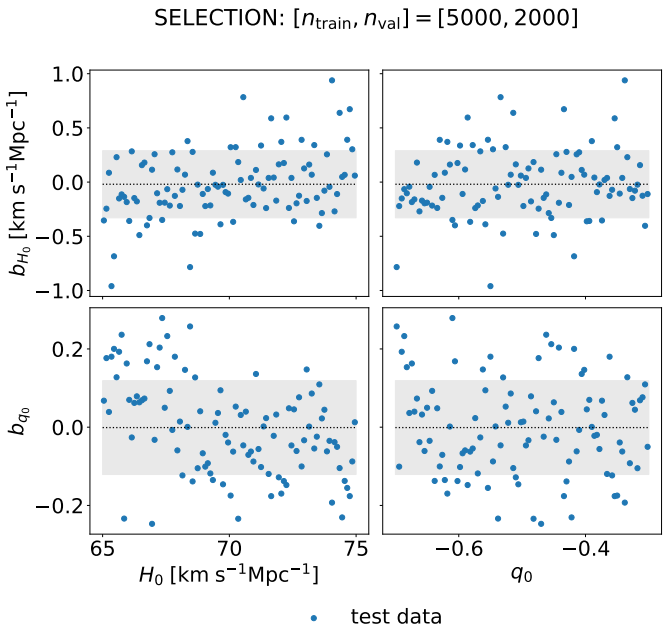

Figure 4.7: Distribution of generative parameters and LFI posterior biases for the GW selection setting. The one-sigma range of the bias is shaded grey. The neural network model used to perform the compression and generate this plot corresponds to the NN parameters combination $[n_{batch}, \alpha, \lambda_1] = [100, 5 \times 10^{-4}, 10^{-4}]$ for $[n_{train}, n_{val}] = [5000, 2000]$.

on the cosmological parameters are consistent with zero, and have standard deviations which are a small fraction of the full posterior uncertainty.

4.5 Conclusions

We have investigated the ability of Likelihood Free Inference (LFI) to estimate the cosmological expansion from GW-selected populations of binary neutron star mergers with EM counterparts. When computing the parameter posterior using traditional Bayesian inference, selection effects must be taken into account through the computation of the expected number of detected sources, $\bar{N}$. This is a computationally expensive (and potentially inaccurate) process, even in approximate forms (Tiwari, 2018; Farr, 2019; Mortlock et al., 2019). As LFI does not explicitly evaluate the posterior, instead building a proxy likelihood using neural density estimator fits to parameter–simulated-dataset pairs, there is no need to calculate $\bar{N}$ when performing LFI. Instead, the selection is naturally built into the simulations on which the method is based.

The goal of this work was to compare the precision and accuracy achievable using LFI to that of traditional Bayesian inference in the presence of selection effects. We note that improvements to the traditional found-injection approach broadening injection-set coverage (through, e.g., designer injection sets covering a range of populations) have the potential to improve the resulting $\bar{N}$ estimates. A quantitative comparison (in terms of precision, accuracy and computational cost, and considering more complete data models) of the LFI approach with improved found-injection methods is strongly motivated by the findings of this proof-of-concept work.

In this work we considered GW selection only; adding EM selection would increase the computational burden, making accounting for selection effects even more expensive. We employed "pre-processed" 100-merger datasets, consisting of noisy estimates of redshift, distance and peculiar velocity for each merger, assuming the distances have already been inferred from GW strains (which can be performed rapidly as in Green and Gair (2021) to yield a fully LFI-based pipeline). Given the high dimensionality of the input data, LFI methods require the data to be compressed to a set of summary statistics. We trained ensembles of regression neural networks for this purpose,

passing their outputs to the density-estimation likelihood-free-inference package PYDELFI to infer the cosmological parameters. Both of these stages require the provision of training data: we have presented results for compression networks trained using $[n_{train}, n_{\rm val}] = [5000, 2000]$ and $[n_{train}, n_{val}] = [500000, 100000]$ populations; in all cases PYDELFI was trained using 2000 simulated populations. Given each population contains 100 mergers, the total number of detected mergers required to train the two setups was 9×10^5 and 6×10^8, respectively.

LFI's precision and accuracy depends sensitively on the compression method's ability to retain salient information about the parameters of interest. We trained a large suite of regression networks (each containing two hidden layers of 128 hidden units) for compression, optimizing the learning rate, batch size and regularization based on PYDELFI's ability to infer H_0 using the networks' outputs. Specifically, we selected the network whose resulting H_0 inference best reproduced the traditional Bayesian "ground truth" (as implemented using PYSTAN) for a set of 100 test datasets, taking the differences between maximum-posterior H_0 estimates for the two methods as our metric.

Testing the method first on datasets in which no GW selection was made, we demonstrated that LFI provides unbiased H_0 estimates when using suitably optimized regression-network data compression. For our optimal combination of training variables, we found a bias (defined as the difference between the maximum-posterior PYSTAN and PYDELFI estimates) on H_0 of $b_{H_0} = 0.021 \pm 0.322\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$: consistent with zero and with a standard deviation a factor of roughly three smaller than the posterior uncertainty on H_0. Marginalizing over this bias would lead to an increase of only 6.45% in the uncertainty on H_0. Adding in GW selection, we find no impact on LFI's performance: LFI is still able to provide unbiased estimates of H_0 in the presence of selection effects. For the best model we obtain $b_{H_0} = -0.02 \pm 0.313\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, which would yield an increase in uncertainty on H_0 of only 5.9% when marginalized over. Increasing the number of samples used to train the compression networks results in LFI posteriors that are statistically indistinguishable from their traditional Bayesian counterparts in mean and variance; however, this comes with a significant increase in computational cost. When processing GW-

selected data, we note a small but significant correlation between the H_0 and q_0 biases and the generative H_0 values. This indicates a different choice of compressor architecture and setup might improve results, but investigating alternative compression methods is left for future work.

As this method is simulation-based, having a trustworthy and sufficient generative model is critical. This analysis has been conducted on simplified mock data, for which we know the underlying model. In the context of real observations, more-realistic simulations, such as those implemented in LALSuite (LIGO Scientific Collaboration, 2018), are needed. As current ground-based interferometers enhance their sensitivity (Abbott et al., 2020), third-generation GW detectors such as Einstein Telescope (Sathyaprakash et al., 2012) and Cosmic Explorer (Abbott et al., 2017) come online, and the BNS sample builds, including instrumental systematics (Sun et al., 2020) and an as-yet elusive model of joint EM-GW selection (e.g. Rosswog et al., 2017; Scolnic et al., 2017; Cowperthwaite et al., 2019; Setzer et al., 2019; Chen, 2020; Mastrogiovanni et al., 2021; Feeney et al., 2021; Raaijmakers et al., 2021) will become ever more important. In this work we have focused on inferring the cosmological parameters only, but complete inference of the population properties of BNS catalogues must include parameters fixed here, such as the merger rate, mass distributions and equation of state (e.g. Abbott et al., 2018; Farrow et al., 2019; Landry et al., 2020; The LIGO Scientific Collaboration et al., 2021; Galaudage et al., 2021; Mastrogiovanni et al., 2021). Extending the analysis to incorporate these parameters is left to future work. Finally, we note that, though we have focused on the inference of the cosmological expansion from GW-selected catalogues of binary neutron star mergers with EM counterparts here, this method can be applied to a broad range of population analyses in the presence of selection effects (The LIGO Scientific Collaboration et al., 2021; Kim, 2021).

The code is provided at `https://github.com/frgerardi/LFIH0_BNS.git`.

4.6 Appendix: Full tables

For completeness, in the following we tabulate the results for all combinations of learning rate, batchsize and regularization explored for both no-selection and selection analyses.

NO SELECTION CASE							
n_{batch}	α	regularizer	b_{H_0} [km s^{-1} Mpc^{-1}]	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [5000, 2000]$							
100	10^{-4}	–	0.369 ± 1.752	0.002 ± 0.098	1.951 ± 0.124	0.948 ± 0.023	165.74%
		$\lambda_2 = 10^{-4}$	-0.002 ± 0.459	0.008 ± 0.089	1.055 ± 0.043	0.95 ± 0.028	15.59%
		$\lambda_2 = 2 \times 10^{-4}$	0.054 ± 0.415	0.008 ± 0.095	1.056 ± 0.05	0.95 ± 0.028	13.95%
		$\lambda_1 = 10^{-4}$	0.024 ± 0.35	-0.003 ± 0.095	1.014 ± 0.045	0.95 ± 0.035	7.64%
		$\lambda_1 = 2 \times 10^{-4}$	0.012 ± 0.398	0.009 ± 0.1	1.038 ± 0.049	0.95 ± 0.037	11.58%
	5×10^{-4}	–	-0.101 ± 1.612	-0.003 ± 0.103	1.855 ± 0.116	0.948 ± 0.031	148.89%
		$\lambda_2 = 10^{-4}$	0.008 ± 0.404	0.012 ± 0.082	1.043 ± 0.039	0.948 ± 0.031	12.24%
		$\lambda_2 = 2 \times 10^{-4}$	-0.003 ± 0.423	0.011 ± 0.084	1.04 ± 0.042	0.948 ± 0.029	12.76%
		$\lambda_1 = 10^{-4}$	-0.002 ± 0.365	0.004 ± 0.098	1.028 ± 0.042	0.952 ± 0.032	9.43%
		$\lambda_1 = 2 \times 10^{-4}$	-0.006 ± 0.415	0.011 ± 0.083	1.027 ± 0.042	0.947 ± 0.03	11.27%
	10^{-3}	–	0.053 ± 0.498	0.005 ± 0.09	1.058 ± 0.037	0.954 ± 0.028	17.55%
		$\lambda_2 = 10^{-4}$	-0.014 ± 0.385	0.009 ± 0.083	1.04 ± 0.045	0.952 ± 0.032	11.32%
		$\lambda_2 = 2 \times 10^{-4}$	-0.012 ± 0.391	0.005 ± 0.09	1.036 ± 0.048	0.949 ± 0.031	11.15%
		$\lambda_1 = 10^{-4}$	0.007 ± 0.352	0.009 ± 0.09	1.024 ± 0.048	0.952 ± 0.038	8.58%
		$\lambda_1 = 2 \times 10^{-4}$	-0.026 ± 0.418	0.011 ± 0.086	1.022 ± 0.048	0.948 ± 0.035	10.94%
500	10^{-4}	–	0.087 ± 2.066	-0.007 ± 0.105	2.047 ± 0.163	0.948 ± 0.025	195.21%
		$\lambda_2 = 10^{-4}$	-0.035 ± 0.41	0.003 ± 0.099	1.049 ± 0.047	0.948 ± 0.032	13.05%
		$\lambda_2 = 2 \times 10^{-4}$	-0.007 ± 0.375	-0.004 ± 0.087	1.04 ± 0.044	0.95 ± 0.04	10.91%
		$\lambda_1 = 10^{-4}$	0.012 ± 0.358	-0.003 ± 0.092	1.003 ± 0.043	0.947 ± 0.036	6.81%
		$\lambda_1 = 2 \times 10^{-4}$	0.01 ± 0.399	0.002 ± 0.096	1.041 ± 0.043	0.948 ± 0.032	11.94%
	5×10^{-4}	–	-0.195 ± 1.807	0.003 ± 0.099	2.068 ± 0.157	0.948 ± 0.022	178.17%
		$\lambda_2 = 10^{-4}$	0.021 ± 0.427	0.003 ± 0.096	1.041 ± 0.042	0.949 ± 0.026	13.01%
		$\lambda_2 = 2 \times 10^{-4}$	-0.027 ± 0.388	0.009 ± 0.099	1.038 ± 0.053	0.953 ± 0.035	11.19%
		$\lambda_1 = 10^{-4}$	0.026 ± 0.328	0.001 ± 0.091	1.018 ± 0.051	0.948 ± 0.026	7.3%
		$\lambda_1 = 2 \times 10^{-4}$	0.015 ± 0.361	-0.002 ± 0.087	1.022 ± 0.043	0.951 ± 0.032	8.73%
	10^{-3}	–	0.593 ± 1.892	0.001 ± 0.101	2.078 ± 0.172	0.951 ± 0.022	184.82%
		$\lambda_2 = 10^{-4}$	0.01 ± 0.413	0.01 ± 0.088	1.052 ± 0.044	0.949 ± 0.034	13.42%
		$\lambda_2 = 2 \times 10^{-4}$	0.003 ± 0.413	0.001 ± 0.084	1.038 ± 0.04	0.947 ± 0.036	12.13%
		$\lambda_1 = 10^{-4}$	0.021 ± 0.322	-0.0 ± 0.087	1.012 ± 0.054	0.943 ± 0.036	6.45%
		$\lambda_1 = 2 \times 10^{-4}$	-0.01 ± 0.362	0.01 ± 0.101	1.03 ± 0.053	0.95 ± 0.041	9.59%

Table 4.6.1: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for all combinations of batchsize, learning rate and regularization in the no-selection case, using $[n_{train}, n_{val}] = [5000, 2000]$.

NO SELECTION CASE							
n_{batch}	α	regularizer	b_{H_0} $[\mathrm{km\,s^{-1}\,Mpc^{-1}}]$	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [500000, 100000]$							
100	10^{-4}	–	-0.063 ± 0.253	0.016 ± 0.065	0.969 ± 0.042	0.945 ± 0.038	0.3%
		$\lambda_2 = 10^{-4}$	-0.053 ± 0.193	0.018 ± 0.065	0.981 ± 0.047	0.951 ± 0.042	0.11%
		$\lambda_2 = 2 \times 10^{-4}$	-0.073 ± 0.193	0.015 ± 0.061	0.979 ± 0.042	0.945 ± 0.038	-0.05%
		$\lambda_1 = 10^{-4}$	-0.073 ± 0.243	0.023 ± 0.062	0.97 ± 0.044	0.944 ± 0.038	0.13%
		$\lambda_1 = 2 \times 10^{-4}$	-0.075 ± 0.254	0.006 ± 0.064	0.975 ± 0.038	0.943 ± 0.034	0.93%
	5×10^{-4}	–	-0.061 ± 0.218	0.014 ± 0.071	0.978 ± 0.048	0.948 ± 0.04	0.35%
		$\lambda_2 = 10^{-4}$	-0.058 ± 0.222	0.015 ± 0.063	0.972 ± 0.045	0.945 ± 0.041	-0.18%
		$\lambda_2 = 2 \times 10^{-4}$	-0.044 ± 0.224	0.016 ± 0.058	0.981 ± 0.049	0.945 ± 0.037	0.79%
		$\lambda_1 = 10^{-4}$	-0.032 ± 0.267	0.009 ± 0.067	0.968 ± 0.045	0.94 ± 0.041	0.57%
		$\lambda_1 = 2 \times 10^{-4}$	-0.072 ± 0.293	0.024 ± 0.062	0.973 ± 0.043	0.947 ± 0.04	1.89%
	10^{-3}	–	-0.058 ± 0.21	0.02 ± 0.058	0.973 ± 0.042	0.945 ± 0.04	-0.35%
		$\lambda_2 = 10^{-4}$	-0.066 ± 0.224	0.024 ± 0.063	0.979 ± 0.044	0.944 ± 0.035	0.54%
		$\lambda_2 = 2 \times 10^{-4}$	-0.039 ± 0.252	0.022 ± 0.063	0.979 ± 0.047	0.946 ± 0.039	1.23%
		$\lambda_1 = 10^{-4}$	-0.074 ± 0.281	0.021 ± 0.062	0.98 ± 0.044	0.947 ± 0.038	2.14%
		$\lambda_1 = 2 \times 10^{-4}$	-0.057 ± 0.3	0.014 ± 0.062	0.974 ± 0.044	0.946 ± 0.039	2.12%
500	10^{-4}	–	-0.034 ± 0.264	0.017 ± 0.065	0.974 ± 0.044	0.948 ± 0.042	1.15%
		$\lambda_2 = 10^{-4}$	-0.043 ± 0.193	0.017 ± 0.066	0.972 ± 0.041	0.944 ± 0.039	-0.77%
		$\lambda_2 = 2 \times 10^{-4}$	-0.047 ± 0.196	0.016 ± 0.059	0.976 ± 0.044	0.945 ± 0.034	-0.37%
		$\lambda_1 = 10^{-4}$	-0.064 ± 0.225	0.012 ± 0.061	0.969 ± 0.041	0.947 ± 0.038	-0.41%
		$\lambda_1 = 2 \times 10^{-4}$	-0.069 ± 0.246	0.022 ± 0.064	0.976 ± 0.044	0.945 ± 0.041	0.88%
	5×10^{-4}	–	-0.052 ± 0.243	0.008 ± 0.06	0.974 ± 0.042	0.946 ± 0.04	0.54%
		$\lambda_2 = 10^{-4}$	-0.064 ± 0.208	0.016 ± 0.057	0.969 ± 0.04	0.942 ± 0.037	-0.79%
		$\lambda_2 = 2 \times 10^{-4}$	-0.053 ± 0.208	0.015 ± 0.061	0.975 ± 0.046	0.944 ± 0.037	-0.15%
		$\lambda_1 = 10^{-4}$	-0.056 ± 0.249	0.022 ± 0.065	0.967 ± 0.039	0.944 ± 0.034	0.0%
		$\lambda_1 = 2 \times 10^{-4}$	-0.057 ± 0.273	0.022 ± 0.07	0.974 ± 0.043	0.946 ± 0.038	1.37%
	10^{-3}	–	-0.056 ± 0.227	0.019 ± 0.056	0.975 ± 0.054	0.944 ± 0.04	0.3%
		$\lambda_2 = 10^{-4}$	-0.062 ± 0.189	0.012 ± 0.06	0.979 ± 0.048	0.946 ± 0.035	-0.22%
		$\lambda_2 = 2 \times 10^{-4}$	-0.044 ± 0.229	0.015 ± 0.066	0.982 ± 0.045	0.946 ± 0.04	0.96%
		$\lambda_1 = 10^{-4}$	-0.076 ± 0.27	0.014 ± 0.063	0.97 ± 0.04	0.946 ± 0.033	0.88%
		$\lambda_1 = 2 \times 10^{-4}$	-0.078 ± 0.285	0.011 ± 0.064	0.974 ± 0.041	0.947 ± 0.037	1.73%

Table 4.6.2: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for all combinations of batchsize, learning rate and regularization in the no-selection case, using $[n_{train}, n_{val}] = [500000, 100000]$.

SELECTION CASE							
n_{batch}	α	regularizer	b_{H_0} $[\mathrm{km\,s^{-1}\,Mpc^{-1}}]$	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
TRAINING and VALIDATION parameters: $[n_{train}, n_{\mathrm{val}}] = [5000, 2000]$							
100	10^{-4}	–	-0.153 ± 1.714	0.014 ± 0.136	1.77 ± 0.139	1.019 ± 0.076	144.02%
		$\lambda_2 = 10^{-4}$	0.043 ± 0.403	0.02 ± 0.123	1.059 ± 0.044	1.012 ± 0.04	13.05%
		$\lambda_2 = 2 \times 10^{-4}$	0.021 ± 0.426	0.016 ± 0.114	1.047 ± 0.037	1.013 ± 0.044	12.74%
		$\lambda_1 = 10^{-4}$	-0.015 ± 0.338	0.014 ± 0.115	1.013 ± 0.039	1.005 ± 0.044	6.53%
		$\lambda_1 = 2 \times 10^{-4}$	-0.029 ± 0.365	0.008 ± 0.119	1.029 ± 0.038	1.008 ± 0.044	8.97%
	5×10^{-4}	–	-0.309 ± 1.657	0.007 ± 0.126	1.838 ± 0.153	1.013 ± 0.032	145.28%
		$\lambda_2 = 10^{-4}$	0.059 ± 0.423	0.009 ± 0.116	1.018 ± 0.043	1.007 ± 0.057	9.95%
		$\lambda_2 = 2 \times 10^{-4}$	-0.015 ± 0.4	0.022 ± 0.118	1.033 ± 0.041	1.009 ± 0.044	10.47%
		$\lambda_1 = 10^{-4}$	-0.02 ± 0.313	-0.001 ± 0.122	1.014 ± 0.041	1.005 ± 0.058	5.9%
		$\lambda_1 = 2 \times 10^{-4}$	0.0 ± 0.393	0.003 ± 0.115	1.036 ± 0.038	1.009 ± 0.042	10.5%
	10^{-3}	–	-0.006 ± 0.526	0.011 ± 0.132	1.055 ± 0.055	1.015 ± 0.089	17.38%
		$\lambda_2 = 10^{-4}$	-0.03 ± 0.408	0.011 ± 0.114	1.013 ± 0.044	1.005 ± 0.056	8.93%
		$\lambda_2 = 2 \times 10^{-4}$	0.017 ± 0.387	0.02 ± 0.121	1.023 ± 0.046	1.009 ± 0.063	9.09%
		$\lambda_1 = 10^{-4}$	0.014 ± 0.357	0.008 ± 0.119	1.018 ± 0.042	1.006 ± 0.051	7.67%
		$\lambda_1 = 2 \times 10^{-4}$	-0.003 ± 0.476	0.0 ± 0.12	1.055 ± 0.048	1.011 ± 0.056	15.32%
500	10^{-4}	–	0.022 ± 1.961	0.005 ± 0.136	2.028 ± 0.176	1.005 ± 0.028	179.37%
		$\lambda_2 = 10^{-4}$	0.011 ± 0.455	0.016 ± 0.119	1.044 ± 0.035	1.009 ± 0.051	13.52%
		$\lambda_2 = 2 \times 10^{-4}$	-0.009 ± 0.417	0.011 ± 0.114	1.056 ± 0.04	1.008 ± 0.048	13.25%
		$\lambda_1 = 10^{-4}$	-0.002 ± 0.334	0.019 ± 0.116	1.025 ± 0.04	1.01 ± 0.049	7.63%
		$\lambda_1 = 2 \times 10^{-4}$	-0.018 ± 0.377	0.013 ± 0.125	1.033 ± 0.037	1.016 ± 0.03	9.66%
	5×10^{-4}	–	0.189 ± 2.115	0.027 ± 0.127	1.97 ± 0.134	1.017 ± 0.042	185.92%
		$\lambda_2 = 10^{-4}$	-0.039 ± 0.453	0.02 ± 0.12	1.048 ± 0.039	1.012 ± 0.056	13.75%
		$\lambda_2 = 2 \times 10^{-4}$	0.011 ± 0.408	0.015 ± 0.119	1.039 ± 0.035	1.01 ± 0.046	11.34%
		$\lambda_1 = 10^{-4}$	0.001 ± 0.313	0.012 ± 0.128	1.025 ± 0.038	1.013 ± 0.036	6.99%
		$\lambda_1 = 2 \times 10^{-4}$	-0.037 ± 0.359	0.014 ± 0.12	1.038 ± 0.044	1.01 ± 0.058	9.62%
	10^{-3}	–	-0.082 ± 1.972	0.016 ± 0.117	2.027 ± 0.173	1.012 ± 0.031	180.05%
		$\lambda_2 = 10^{-4}$	-0.006 ± 0.451	0.011 ± 0.119	1.046 ± 0.036	1.005 ± 0.062	13.57%
		$\lambda_2 = 2 \times 10^{-4}$	-0.027 ± 0.371	0.017 ± 0.124	1.046 ± 0.042	1.005 ± 0.058	10.71%
		$\lambda_1 = 10^{-4}$	0.051 ± 0.329	0.011 ± 0.137	1.019 ± 0.053	1.011 ± 0.058	6.91%
		$\lambda_1 = 2 \times 10^{-4}$	0.004 ± 0.348	0.011 ± 0.113	1.027 ± 0.038	1.003 ± 0.05	8.2%

Table 4.6.3: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for all combinations of batchsize, learning rate and regularization in the selection case, using $[n_{train}, n_{\mathrm{val}}] = [5000, 2000]$.

n_{batch}	α	regularizer	b_{H_0} $[\mathrm{km\,s^{-1}\,Mpc^{-1}}]$	b_{q_0}	$f_\sigma^{H_0}$	$f_\sigma^{q_0}$	$\%\hat{\sigma}_{incr}^{H_0}$
SELECTION CASE							
TRAINING and VALIDATION parameters: $[n_{train}, n_{val}] = [500000, 100000]$							
100	10^{-4}	–	-0.033 ± 0.278	0.023 ± 0.082	0.97 ± 0.037	0.999 ± 0.045	0.78%
		$\lambda_2 = 10^{-4}$	-0.032 ± 0.184	0.022 ± 0.092	0.976 ± 0.031	1.006 ± 0.036	-0.73%
		$\lambda_2 = 2 \times 10^{-4}$	-0.019 ± 0.195	0.017 ± 0.085	0.981 ± 0.036	1.002 ± 0.037	-0.04%
		$\lambda_1 = 10^{-4}$	-0.025 ± 0.186	0.021 ± 0.085	0.978 ± 0.033	1.003 ± 0.039	-0.55%
		$\lambda_1 = 2 \times 10^{-4}$	-0.044 ± 0.214	0.018 ± 0.087	0.984 ± 0.036	1.005 ± 0.042	0.57%
	5×10^{-4}	–	0.01 ± 0.207	0.02 ± 0.087	0.968 ± 0.038	1.001 ± 0.04	-1.12%
		$\lambda_2 = 10^{-4}$	-0.033 ± 0.177	0.02 ± 0.092	0.979 ± 0.039	1.003 ± 0.043	-0.56%
		$\lambda_2 = 2 \times 10^{-4}$	-0.026 ± 0.199	0.015 ± 0.088	0.979 ± 0.037	1.002 ± 0.04	-0.21%
		$\lambda_1 = 10^{-4}$	-0.028 ± 0.198	0.019 ± 0.081	0.988 ± 0.037	1.001 ± 0.042	0.64%
		$\lambda_1 = 2 \times 10^{-4}$	-0.047 ± 0.269	0.018 ± 0.084	0.989 ± 0.034	1.0 ± 0.045	2.37%
	10^{-3}	–	0.0 ± 0.183	0.026 ± 0.091	0.965 ± 0.03	1.004 ± 0.038	-1.88%
		$\lambda_2 = 10^{-4}$	-0.007 ± 0.184	0.014 ± 0.088	0.98 ± 0.034	1.005 ± 0.044	-0.35%
		$\lambda_2 = 2 \times 10^{-4}$	-0.015 ± 0.193	0.015 ± 0.093	0.982 ± 0.037	1.004 ± 0.044	-0.0%
		$\lambda_1 = 10^{-4}$	-0.053 ± 0.242	0.015 ± 0.087	0.996 ± 0.033	1.001 ± 0.045	2.39%
		$\lambda_1 = 2 \times 10^{-4}$	-0.031 ± 0.263	0.015 ± 0.095	0.991 ± 0.036	1.001 ± 0.048	2.44%
500	10^{-4}	–	-0.037 ± 0.267	0.022 ± 0.084	0.98 ± 0.041	0.998 ± 0.036	1.42%
		$\lambda_2 = 10^{-4}$	-0.038 ± 0.199	0.028 ± 0.109	0.976 ± 0.034	1.01 ± 0.035	-0.51%
		$\lambda_2 = 2 \times 10^{-4}$	-0.02 ± 0.194	0.021 ± 0.095	0.971 ± 0.034	1.005 ± 0.036	-1.05%
		$\lambda_1 = 10^{-4}$	-0.022 ± 0.178	0.015 ± 0.093	0.978 ± 0.036	1.003 ± 0.039	-0.7%
		$\lambda_1 = 2 \times 10^{-4}$	-0.025 ± 0.186	0.012 ± 0.089	0.979 ± 0.036	1.002 ± 0.035	-0.47%
	5×10^{-4}	–	-0.045 ± 0.277	0.019 ± 0.09	0.982 ± 0.039	1.002 ± 0.037	1.83%
		$\lambda_2 = 10^{-4}$	-0.013 ± 0.18	0.019 ± 0.086	0.977 ± 0.035	1.006 ± 0.038	-0.68%
		$\lambda_2 = 2 \times 10^{-4}$	-0.02 ± 0.182	0.014 ± 0.087	0.981 ± 0.032	1.005 ± 0.039	-0.32%
		$\lambda_1 = 10^{-4}$	-0.03 ± 0.196	0.018 ± 0.079	0.982 ± 0.032	1.001 ± 0.036	0.04%
		$\lambda_1 = 2 \times 10^{-4}$	-0.047 ± 0.233	0.012 ± 0.089	0.987 ± 0.038	1.004 ± 0.036	1.33%
	10^{-3}	–	-0.013 ± 0.22	0.021 ± 0.089	0.962 ± 0.042	1.003 ± 0.044	-1.43%
		$\lambda_2 = 10^{-4}$	-0.006 ± 0.201	0.015 ± 0.086	0.98 ± 0.033	1.004 ± 0.041	-0.07%
		$\lambda_2 = 2 \times 10^{-4}$	-0.01 ± 0.199	0.021 ± 0.083	0.979 ± 0.043	1.003 ± 0.042	-0.18%
		$\lambda_1 = 10^{-4}$	-0.026 ± 0.212	0.015 ± 0.084	0.986 ± 0.037	1.0 ± 0.044	0.77%
		$\lambda_1 = 2 \times 10^{-4}$	-0.036 ± 0.25	0.012 ± 0.09	0.993 ± 0.037	1.005 ± 0.043	2.28%

Table 4.6.4: Means and standard deviations for the biases b_{H_0,q_0}, posterior-width ratios f_{H_0,q_0} and percentage increase in H_0 uncertainty for all combinations of batchsize, learning rate and regularization in the selection case, using $[n_{train}, n_{val}] = [500000, 100000]$.

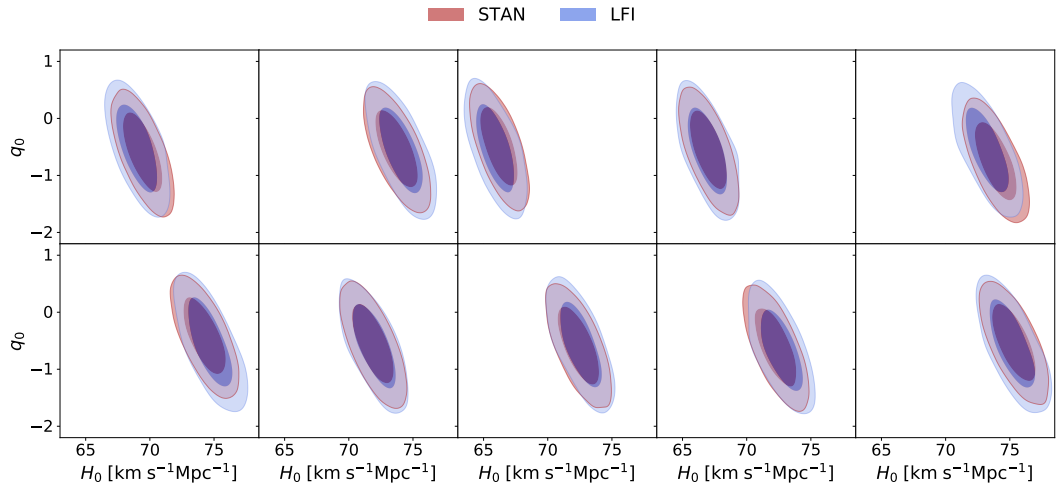

Figure 4.6.1: This plot shows how the LFI (blue) and traditional Bayesian sampling (red) posterior contours compare for ten random BNS catalogues (ten subplots) generated with GW selection applied. In particular, for each subplot the x-axis and y-axis are the Hubble constant H_0 and the deceleration parameter q_0 respectively. These ten catalogues are part of the hundred used to build the statistical analysis presented in Sect. 4.4.2 and their contour plots are examples of the excellent agreement between the LFI and traditional frameworks on a statistical level.

5 Direct cosmological inference from 3D Lyman-α correlations

When performing cosmological inference, standard analyses of the Lyman-α (Lyα) three-dimensional correlation functions only consider the information carried by the distinct peak produced by baryon acoustic oscillations (BAO). In this work, we address whether this compression is sufficient to capture all the relevant cosmological information carried by these functions. We do this by performing a direct fit to the full shape, including all physical scales without compression, of synthetic Lyα auto-correlation functions and cross-correlations with quasars at effective redshift $z_{\rm eff} = 2.3$, assuming a DESI-like survey, and providing a comparison to the classic method applied to the same dataset. Our approach leads to a 3.5% constraint on the matter density $\Omega_{\rm M}$, which is about three to four times better than what BAO alone can probe. The growth term $f\sigma_8(z_{\rm eff})$ is constrained to the 10% level, and the spectral index $n_{\rm s}$ to $\sim 3-4\%$. We demonstrate that the extra information resulting from our 'direct fit' approach, except for the $n_{\rm s}$ constraint, can be traced back to the Alcock-Paczyński effect and redshift space distortion information.

The work presented in this chapter has been published, see Gerardi et al. (2022)

5.1 Introduction

Over the last couple of decades, after the discovery of the accelerated expansion of the Universe (Riess et al., 1998; Perlmutter et al., 1999), cosmology has focused on investigating the properties of dark energy. Among the multiple probes used to place a contraint on the parameters of the current ΛCDM model, there is the baryon acoustic oscillation (BAO) scale over different tracers[1]

[1]For a BAO distance ladder plot https://www.sdss.org/science/cosmology-results-from-eboss/.

of the matter density field (Eisenstein et al., 2005; Cole et al., 2005). Measurements of this standard ruler over a range of redshifts place a constraint on the expansion history (Seo and Eisenstein, 2003).

Complementary to low-redshift galaxies ($z \lesssim 1$), the Lyman-α (Lyα) forest is a tracer of the intergalactic medium (IGM) that probes the cosmic expansion via BAO at higher redshifts, as first proposed by McDonald and Eisenstein (2007). The Lyα forest is a sequence of absorption lines in high-redshift quasar (QSO) spectra, caused by the neutral hydrogen distributed along the line of sight, between the quasar and the observer. The first BAO detection from the Lyα auto-correlation function and from its cross-correlation with QSOs was in the Baryon Oscillation Spectroscopic Survey (BOSS) DR9 data (Busca et al., 2013; Slosar et al., 2013; Kirkby et al., 2013) and DR11 data (Font-Ribera et al., 2014), respectively.

BAO produce a distinct feature in the correlation functions, which we wish to measure and use to probe cosmology, in a robust and model-independent way. When performing cosmological inference, a standard method, as applied in BOSS and eBOSS (du Mas des Bourboux et al., 2020) analyses of the Lyα three-dimensional correlation functions, relies on splitting them into a *peak* and a *smooth* component and only considers the information carried by the BAO feature.

Recently, Cuceu et al. (2021) (C21 hereafter) demonstrated that further cosmological information can be obtained from the broadband component using the Alcock-Paczyński (AP) effect (Alcock and Paczynski, 1979). When computing the 3D correlation functions from observations, as a standard approach, we change from angular and redshift separations to comoving coordinates, based on the assumption of a fiducial cosmological model. In particular, if the latter differs from the true underlying cosmology, then the AP effect will appear as an apparent anisotropy in the correlation functions. Another source of anisotropy is redshift space distortions (RSD), which are induced by peculiar velocities and hence carry extra information. However, measuring redshift space distortions for the Lyα auto-correlation alone is not informative about the growth rate of structure because of its degeneracy with an unknown velocity divergence bias (Seljak, 2012).

For this reason, C21 jointly employed the Lyα auto- and cross-correlation with quasars to explore the potential of measuring the linear growth of structure.

All physical scales of the 3D Lyα correlation functions, beyond the BAO peak, carry information about the underlying cosmology. Throughout this work we will refer to the sum of all of these scales, with no compression, as the full shape of these functions. Both the Ly$\alpha\times$Lyα auto- and Ly$\alpha\times$QSO cross-correlation functions can be directly used to perform cosmological inference. The work of C21 motivates a further investigation, assessing whether or not the compressed analysis based on BAO, AP and RSD successfully captures all cosmological information from the correlation functions of interest.

The same point is relevant also in the field of galaxy clustering, where the compressed standard approach extracts cosmological information from BAO, AP and RSD (Alam et al., 2017, 2021). Over the past few years, advancements in perturbation theory computations boosted the interest in fitting the observed two-point statistics and directly inferring cosmological parameters without compression (Tröster et al., 2020; Ivanov et al., 2020; d'Amico et al., 2020). In particular, Brieden et al. (2021) carried out such an investigation, identifying from where in the data vector additional information originates and extending the classic approach by introducing an extra physical parameter.

In this work, we aim to address whether or not compression is a suitable approach in the field of Lyα forest cosmology, by performing a direct fit to the full shape of synthetic Ly$\alpha\times$Lyα auto- and Ly$\alpha\times$QSO cross-correlation functions, and by subsequently comparing the constraints with those obtained using the standard approach. In real data, Lyα correlations have contaminants that affect the amount of cosmological information extractable from a given scale. We discuss here an optimistic scenario without contaminants, and leave for future work a detailed study of their impact on the cosmological constraints.

The paper is structured as follows. We start in Sect. 5.2 by outlining the methodology and explaining the inference framework we use. We proceed to apply this method to noiseless synthetic

Parameter	Fiducial	Prior	68% limits
$H_0[\mathrm{km/(s \times Mpc)}]$	67.31	$\mathcal{U}(40, 100)$	$67.69^{+5.5}_{-3.16}$
Ω_M	0.3144	$\mathcal{U}(0.01, 0.99)$	0.318 ± 0.011
$\Omega_\mathrm{B}h^2$	0.02222	$\mathcal{U}(0.01, 0.05)$	$0.0229^{+0.0064}_{-0.0038}$
A_s	$2.196 \cdot 10^{-9}$	$\mathcal{U}_{\log(10^{10}\mathrm{A_s})}(0.5, 6)$	$\left(2.06^{+0.42}_{-0.46}\right) \cdot 10^{-9}$
n_s	0.9655	$\mathcal{U}(0.8, 1.2)$	$0.958^{+0.025}_{-0.035}$
$b_{\mathrm{Ly}\alpha}$	-0.117	$\mathcal{U}_{\log(-\mathrm{b_{Ly\alpha}})}(-2, 0)$	$-0.111^{+0.011}_{-0.012}$
$\beta_{\mathrm{Ly}\alpha}$	1.67	$\mathcal{U}(0, 5)$	1.67 ± 0.03
b_QSO	3.8	$\mathcal{U}_{\log(\mathrm{b_{QSO}})}(-2, 1.3)$	$3.61^{+0.47}_{-0.32}$
$\sigma_\mathrm{v}(\mathrm{M}pc/h)$	6.86	$\mathcal{U}(0, 15)$	$6.75^{+0.64}_{-0.55}$

Table 5.2.1: Full set of sampled parameters, alongside with the fiducial values used to compute the synthetic correlations and the uniform ($\mathcal{U}$) priors adopted for the sampling procedure. When sampling in logarithmic space we add a 'log' subscript to $\mathcal{U}$. In the last column, we provide the one-dimensional marginals (68% c.l.) for all the parameters sampled, where for any asymmetric posteriors we report the posterior maximum with lower and upper 68% limits.

correlation functions and present the forecasts in Sect. 5.3. We finally compare our main results to the compressed analyses in Sect. 5.4 and draw our conclusions in Sect. 5.5.

5.2 Method

We use the full shape of the 3D Lyα correlation functions to directly infer cosmological parameters, without the usual compression methods. Based on the modelling of C21 and with the aid of CAMB (Lewis et al., 2000; Howlett et al., 2012), we construct a likelihood for the COBAYA framework (Torrado and Lewis, 2021; Torrado and Lewis, 2019), and we perform Markov chain Monte Carlo (MCMC) sampling (Lewis and Bridle, 2002; Neal, 2005; Lewis, 2013) using the Ly$\alpha\times$Lyα auto- and Ly$\alpha\times$QSO cross-correlations.

In this section we outline the key features of the method. In Sect. 5.2.1 we describe how the synthetic correlation data was generated, and focus on the modelling in Sect. 5.2.2. In particular, throughout the analysis, we use the code VEGA[2] (Cuceu et al., 2020), which is based on the code used in eBOSS DR16 (du Mas des Bourboux et al., 2020) for fitting and modelling the Lyα correlation functions. We finally motivate the choice of the sampled parameter space and describe the likelihood in Sect. 5.2.3.

[2] https://github.com/andreicuceu/vega

5.2.1 Synthetic data vector and covariance

In this work, we focus on idealised 3D Lyα synthetic correlations in flat ΛCDM, without contaminants. However, we do include the distortion due to quasar continuum fitting. The latter filters out information and 'distorts' the true correlation function (Bautista et al., 2017; du Mas des Bourboux et al., 2017). Our synthetic data was generated using the framework of C21, and is given by an uncontaminated model based on the best fit of eBOSS DR16 (see Tab. 5.2.1). We did not add noise to the data vector, as we are only interested in forecasting. We used covariance matrices based on DESI mocks similar to those used in Youles et al. (2022). These mocks were created with the CoLoRe (Ramírez-Pérez et al., 2022) and LyaCoLoRe (Farr et al., 2020) packages, covering 14000 sq. degrees with a target density of ~ 50 QSOs / sq. degree (DESI Collaboration et al., 2016). The covariance was computed using the community package `picca` (du Mas des Bourboux et al., 2020). In this analysis, we limit ourselves to linear scales, assuming $r_{\rm min} = 30 {\rm h}^{-1}{\rm Mpc}$, up to $r_{\rm max} = 180 {\rm h}^{-1}{\rm Mpc}$. The effective redshift of the correlation functions is $z_{\rm eff} = 2.3$.

5.2.2 Modelling

To infer cosmology from these synthetic correlations, we first need a theory to model the data given any cosmology $\boldsymbol{p_{\rm C}}$. The theoretical 3D Lyα correlation functions are computed from the isotropic matter power spectrum $P(k)$ and then compared against data to evaluate the likelihood.

When modelling and fitting Lyα correlations, we must match the coordinate grid for the theoretical correlation $\boldsymbol{\xi}$ with the grid of the data. When measuring the 3D Lyα correlation functions from observations, we change from angular $\Delta\theta$ and redshift Δz separations to a set of comoving coordinates $(r_{\parallel}, r_{\perp})$, respectively defined along and across the line of sight. This is motivated by the fact that both the radial comoving distance $D_{\rm C}(z) = c\int_0^z dz/H(z)$ and the comoving angular diameter distance $D_{\rm M}(z)$ are redshift dependent, where c is the speed of light and $H(z)$ the Hubble parameter. For this reason, we wish to refer instead to a set of comoving coordinates. Given two locations at redshift z_i and z_j separated by an angle $\Delta\theta$, these are defined as

$$r_{\parallel} = [D_{\mathrm{C,fid}}(z_i) - D_{\mathrm{C,fid}}(z_j)] \cos \frac{\Delta\theta}{2} \; ; \tag{5.1}$$

$$r_{\perp} = [D_{\mathrm{M,fid}}(z_i) + D_{\mathrm{M,fid}}(z_j)] \sin \frac{\Delta\theta}{2} \; , \tag{5.2}$$

where both D_{C} and D_{M} are computed using an assumed fiducial cosmology (fid subscript). In our case, the fiducial cosmology coincides with the cosmological model that was used to generate the data vector. Given that the sampled cosmology that generated the theoretical $\boldsymbol{\xi}$ can be different from the fiducial one, we need to match coordinate grids of data and $\boldsymbol{\xi}$ by rescaling at each sampling step the coordinates of the correlation via

$$q_{\parallel} = D_{\mathrm{H}}(z_{\mathrm{eff}}) / D_{\mathrm{H}}^{\mathrm{fid}}(z_{\mathrm{eff}}) \; ; \tag{5.3}$$

$$q_{\perp} = D_{\mathrm{M}}(z_{\mathrm{eff}}) / D_{\mathrm{M}}^{\mathrm{fid}}(z_{\mathrm{eff}}) \; , \tag{5.4}$$

where $D_{\mathrm{H}}(z) = c/H(z)$, such that $r'_{\parallel,\perp} = q_{\parallel,\perp} r_{\parallel,\perp}$.

In modelling the Lyα correlation functions of interest we follow Eq. (27) of du Mas des Bourboux et al. (2020), adopting the prescriptions of C21. For any cosmology $\boldsymbol{p}_{\mathbf{C}}$, the power spectra of the tracers are computed from the isotropic linear matter power spectrum $P(k, z)$ as

$$P_{\mathrm{Ly}\alpha}(k, \mu_k, z) = b_{\mathrm{Ly}\alpha}^2 \left(1 + \beta_{\mathrm{Ly}\alpha} \mu_k^2\right)^2 F_{\mathrm{nl,Ly}\alpha}^2(k, \mu_k) P(k, z) \; ; \tag{5.5}$$

$$P_{\times}(k, \mu_k, z) = b_{\mathrm{Ly}\alpha} \left(1 + \beta_{\mathrm{Ly}\alpha} \mu_k^2\right) \times \left(b_{\mathrm{QSO}} + f(z) \mu_k^2\right) F_{\mathrm{nl,QSO}}(k_{\parallel}) P(k, z) \; , \tag{5.6}$$

with $f(z) = \frac{d \ln D}{d \ln a}$ being the logarithmic growth rate and $\mu_k = k_{\parallel}/k$, where k and $k_{\parallel}$ are the modulus of the wave vector and its projection along the line of sight respectively. Focusing first on the Ly$\alpha\times$Lyα power spectrum in Eq. (5.5), we identify the Lyα forest linear bias, $b_{\mathrm{Ly}\alpha}$, and its RSD term, $\beta_{\mathrm{Ly}\alpha} = \frac{b_{\eta,\mathrm{Ly}\alpha} f(z)}{b_{\mathrm{Ly}\alpha}}$, where $b_{\eta,\mathrm{Ly}\alpha}$ is an extra unknown bias, the velocity divergence

bias. The choice of using $\beta_{\mathrm{Ly}\alpha}$ in the RSD term comes from the fact that $b_{\eta,\mathrm{Ly}\alpha}$ is fully degenerate with the growth rate $f(z)$. We treat both $b_{\mathrm{Ly}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$ as nuisance parameters and marginalize over them. The $F_{\mathrm{nl,Ly}\alpha}$ term encodes the non-linear corrections according to the model of Arinyo-i-Prats et al. (2015). The parameters involved in this model are kept constant for simplicity, but in principle they should also be varied. However, this should not have a major impact in our analysis, as we are restricting the analysis to linear scales ($r_{\mathrm{min}} = 30\mathrm{h}^{-1}\mathrm{Mpc}$). Moving on to the Ly$\alpha\times$QSO power spectrum $P_\times$ in Eq. (5.6), b_{QSO} is the quasar linear bias, another nuisance parameter. In contrast to the Lyα RSD term, the QSO RSD term is instead simply $f(z)$ as by definition $b_{\eta,\mathrm{QSO}} = 1$. Following du Mas des Bourboux et al. (2020), we model the impact of redshift errors and non-linear peculiar velocities of quasars with a damping term $F_{\mathrm{nl,QSO}}(k_\parallel)$. We use a Lorentzian function

$$F_{\mathrm{nl,QSO}}(k_\parallel) = \sqrt{\left[1 + \left(k_\parallel \sigma_{\mathrm{v}}\right)^2\right]^{-1}}\,, \tag{5.7}$$

with a free parameter σ_{v}, which represents the velocity dispersion and is an extra nuisance parameter.

At each sampling step then, theoretical correlation functions $\boldsymbol{\xi}$ are computed in VEGA. At its core, VEGA decomposes the power spectrum into multipoles, transforms them into correlation function multipoles using the FFTLog algorithm (Hamilton, 2000) and finally reconstructs the two-dimensional correlation function.

5.2.3 Parameter space and likelihood

The BAO feature is able to constrain $\alpha_\parallel = D_{\mathrm{H}}(z) r_{\mathrm{d}}^{\mathrm{fid}} / D_{\mathrm{H}}^{\mathrm{fid}}(z) r_{\mathrm{d}}$ along the line of sight and $\alpha_\perp = D_{\mathrm{M}}(z) r_{\mathrm{d}}^{\mathrm{fid}} / D_{\mathrm{M}}^{\mathrm{fid}} r_{\mathrm{d}}$ in the transverse direction, where r_{d} is the sound horizon at the drag epoch. Since in flat ΛCDM, at low redshifts, the Hubble parameter $H(z)$ can be expressed as a function of H_0 and Ω_{M}, $\alpha_\parallel$ and $\alpha_\perp$ will ultimately place a constraint on $\{H_0 r_{\mathrm{d}}, \Omega_{\mathrm{M}}\}$. Additionally, r_{d} can be numerically approximated as a function of $\Omega_\nu h^2$, $\Omega_{\mathrm{M}} h^2$ and $\Omega_{\mathrm{B}} h^2$ (Aubourg et al., 2015), which are the neutrino, matter and baryon densities respectively, all evaluated at red-

shift $z = 0$ by definition. This further motivates the choice of sampling $\{H_0, \Omega_\mathrm{M}, \Omega_\mathrm{B}h^2\}$, where $\Omega_\nu h^2$ is constant for a given choice of the neutrino mass. Extra information on Ω_M also comes from the AP effect (Alcock and Paczynski, 1979)

$$F_\mathrm{AP} = \frac{a_\perp}{a_\parallel} = \frac{D_\mathrm{M}(z_\mathrm{eff})/D_\mathrm{M}^\mathrm{fid}(z_\mathrm{eff})}{D_\mathrm{H}(z_\mathrm{eff})/D_\mathrm{H}^\mathrm{fid}(z_\mathrm{eff})} = \frac{[D_\mathrm{M}(z_\mathrm{eff})H(z_\mathrm{eff})]}{[D_\mathrm{M}(z_\mathrm{eff})H(z_\mathrm{eff})]_\mathrm{fid}}, \tag{5.8}$$

which is an apparent anisotropy present if the sampled cosmology differs from the fiducial one. For the same assumptions as before, the AP parameter will only be a function of Ω_M. As we fit the full shape of the correlation functions directly and the amplitude of primordial fluctuations A_s and their spectral index n_s affect the functional form of $\boldsymbol{\xi}$, we will sample the full set of parameters $\boldsymbol{p_\mathrm{C}} = \{H_0, \Omega_\mathrm{M}, \Omega_\mathrm{B}h^2, A_\mathrm{s}, n_\mathrm{s}\}$. On the other hand, $\{b_{\mathrm{Ly}\alpha}, b_\mathrm{QSO}, \beta_{\mathrm{Ly}\alpha}, \sigma_\mathrm{v}\}$ are treated as nuisance parameters $\boldsymbol{p_\mathrm{A}}$ to marginalize over.

For all these parameters we choose uniform priors, which are listed in Tab. 5.2.1. As is common, A_s is sampled in logarithmic space, and we made the choice of doing the same with the two linear biases because they are degenerate with A_s and span over several orders of magnitude.

In this work, we assume a Gaussian likelihood, which is also computed using VEGA. A likelihood evaluation, via CAMB, first computes the comoving distances, to calculate $q_\parallel$ and $q_\perp$, and then the isotropic linear matter power spectrum, along with r_d, the growth rate at z_eff and $f\sigma_8(z_\mathrm{eff})$. Then, VEGA computes the correlation functions, based on the modelling description in Sect. 5.2.2, and the χ^2 value.

5.3 Results

In this section we present the forecasts produced using the method outlined in Sect. 5.2 on 3D Ly$\alpha\times$Lyα and Ly$\alpha\times$QSO simplified synthetic correlation functions. We sample over the cosmological parameters $\boldsymbol{p_\mathrm{C}} = \{H_0, \Omega_\mathrm{M}, \Omega_\mathrm{B}h^2, A_\mathrm{s}, n_\mathrm{s}\}$, marginalizing over the astrophysical model parameters $\boldsymbol{p_\mathrm{A}} = \{b_{\mathrm{Ly}\alpha}, b_\mathrm{QSO}, \beta_{\mathrm{Ly}\alpha}, \sigma_\mathrm{v}\}$. The fiducial values of these parameters, along with the priors, are listed in Tab. 5.2.1.

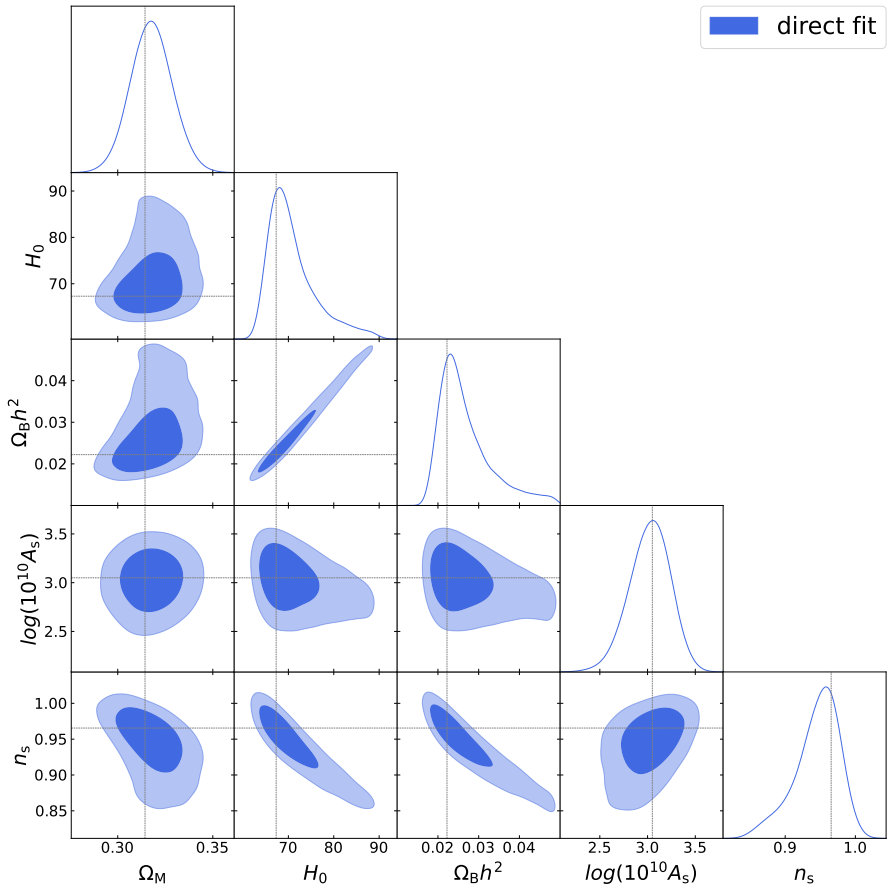

Figure 5.3.1: Triangle plot of the cosmological parameters of interest $\{H_0, \Omega_{\rm M}, \Omega_{\rm B}h^2, A_{\rm s}, n_{\rm s}\}$, marginalizing over the nuisance parameters $\boldsymbol{p}_{\mathbf{A}}$. The blue contours refer to the results obtained performing the inference using the method outlined in Sect. 5.2, which we denote as 'direct fit'. The grey dashed lines mark the fiducial values used to generate data.

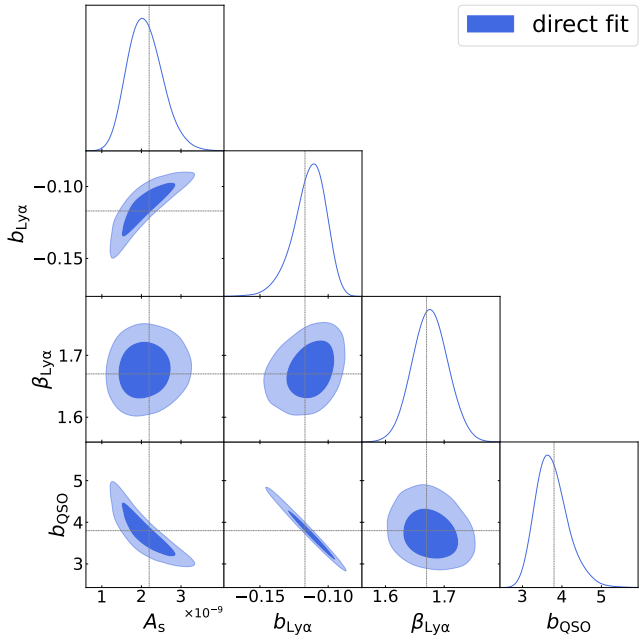

Figure 5.3.2: Correlations among the nuisance parameters $\{b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}\}$ and A_{s}, where 'direct fit' refers to the inference method described in Sect. 5.2. The grey dashed lines mark the fiducial values used to generate the data.

In Fig. 5.3.1 we show the results for $\{H_0, \Omega_{\mathrm{M}}, \Omega_{\mathrm{B}}h^2, A_{\mathrm{s}}, n_{\mathrm{s}}\}$ using the noiseless mock data vector. In the last column of Tab. 5.2.1 we list the one-dimensional marginal constraints for the full set of sampled parameters.

From Fig. 5.3.1, it can be seen that we do recover the true values (shown in grey dashed lines) of cosmological parameters $\boldsymbol{p}_{\mathbf{C}}$ well within 1σ (Tab. 5.2.1). The analysis provides a 3.5% constraint on the matter density Ω_{M}. Clear degeneracies are present between H_0 and $\Omega_{\mathrm{B}}h^2$. As previously mentioned, the baryon acoustic oscillation peak measures the product $H_0 r_{\mathrm{d}}$, which can be expressed as a function of H_0, $\Omega_{\mathrm{B}}h^2$ and Ω_{M}. Given that we have a good measurement of Ω_{M} from the AP information, the remaining degeneracy is between H_0 and $\Omega_{\mathrm{B}}h^2$: if there were no other information on either one of them, these two parameters would be fully degenerate. The fact that both H_0 and $\Omega_{\mathrm{B}}h^2$ are strongly correlated with the spectral index n_{s} could hint that the turnover of the power spectrum is the feature partially breaking the degeneracy. Despite this correlation, we are able to place a constraint on the spectral index, namely $n_s = 0.958^{+0.025}_{-0.035}$. We

obtain a 21% constraint on the amplitude of fluctuations $A_{\rm s}$, with a corresponding constraint on the amplitude of linear matter fluctuations in spheres of $8\,{\rm h}^{-1}{\rm Mpc}$ of $\sigma_8(z_{\rm eff}) = 0.317 \pm 0.032$. We will further analyze the constraining power of our analysis against the state-of-the-art results later in Sect. 5.4.1.

In Fig. 5.3.2 we show the strong correlation among the linear biases, $b_{{\rm Ly}\alpha}$ and $b_{\rm QSO}$, and $A_{\rm s}$, which is expected given the functional form of Eqs. (5.5-5.6). The Lyα auto-correlation alone would not be able to place a constraint on $A_{\rm s}$ since, for $\mu = 0$, we would measure the combination $A_{\rm s} b^2_{{\rm Ly}\alpha}$ only, whereas its anisotropy would provide a measurement of $\beta_{{\rm Ly}\alpha}$. On the other hand, the transverse mode of the cross-power spectrum (Eq. 5.6), combined with the auto-correlation, measures the combination of $A_{\rm s} b^2_{{\rm Ly}\alpha}$ and $b_{\rm QSO}/b_{{\rm Ly}\alpha}$, while through the RSD term we are able to constrain $A_{\rm s}$ (or $f\sigma_8$ in compressed analyses).

5.4 Discussion

In Sect. 5.4.1 we provide a direct comparison of the forecasts presented in Sect. 5.3 and the literature. In particular, we will focus on a comparison with the results on the same synthetic data obtained using the standard BOSS and eBOSS analysis first, as well as the C21 approach. Our goal is to understand whether the compressed analyses successfully capture the cosmological information carried by the 3D Lyα correlation functions and discuss which components of the data are the most informative. In Sect. 5.4.2 we discuss how results change when marginalizing over the growth of structure. This is instructive to further understand from where extra information originates.

5.4.1 Cosmological information

As mentioned above, in flat ΛCDM the BAO scale, along and across the line of sight, identifies a banana-shaped degeneracy in the $[H_0 r_{\rm d}, \Omega_{\rm M}]$ plane. This justifies that any comparison among methods which have the BAO as a primary feature should necessarily happen in this plane. On the other hand, the Lyα-QSO cross-correlation can in principle measure $f\sigma_8(z_{\rm eff})$ because of the

Parameter	BAO	BAO+AP+RSD	direct fit
$\Omega_{\rm M}$	12%	4%	3.5 %
$H_0 r_{\rm d}/c$	4.5%	1.65%	1.43 %
$f\sigma_8(z_{\rm eff})$	–	12.5%	10.4%

Table 5.4.1: Constraining power of our method ('direct fit') on the listed parameters, against those from the standard analysis (BAO) and the one of C21 (BAO+AP+RSD).

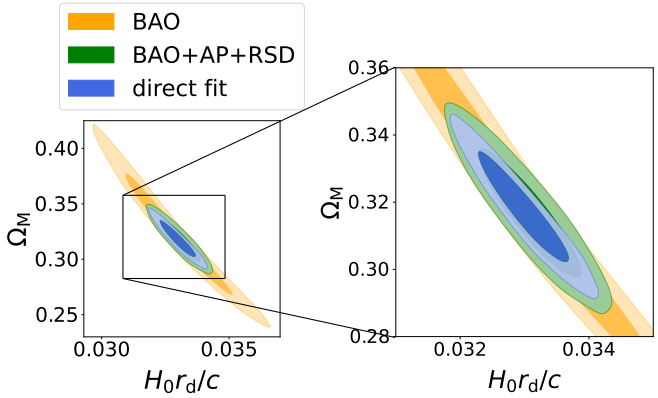

Figure 5.4.1: Two-dimensional contour plots of $\{H_0 r_{\rm d}/c, \Omega_{\rm M}\}$, comparing our method ('direct fit') in blue against standard BOSS and eBOSS analysis ('BAO') in orange and C21 ('BAO+AP+RSD') in green. On the right, there is a zoom-in to further highlight the differences among the 'direct fit' and 'BAO+AP+RSD' methods.

functional form of Eq. (5.6). However, because of the degeneracy with the linear biases, the combination with the Lyα auto-correlation is needed. In what follows we will focus on a comparison based on the derived parameters $\boldsymbol{p_{\rm d}} = \{H_0 r_{\rm d}/c, f\sigma_8(z_{\rm eff})\}$ and $\Omega_{\rm M}$. In particular, in Fig. 5.4.1 we plot the two-dimensional contours of $\{H_0 r_{\rm d}/c, \Omega_{\rm M}\}$ and in Fig. 5.4.2 the one-dimensional marginal of $f\sigma_8(z_{\rm eff})$ for the methods we want to compare.

As discussed above, standard BOSS and eBOSS analyses focus on the peak component of the 3D Lyα correlation functions only. For this reason, we will refer to this approach as 'BAO' for simplicity. We run this analysis using our noiseless mock data, and the most important result is shown in Fig. 5.4.1. This approach provides $\Omega_{\rm M} = 0.32 \pm 0.04$, $H_0 r_{\rm d}/c = 0.0329 \pm 0.0015$, with a constraining power on $H_0 r_{\rm d}/c$ of 4.5%, a factor of three worse compared to our direct fit (summary in Tab. 5.4.1).

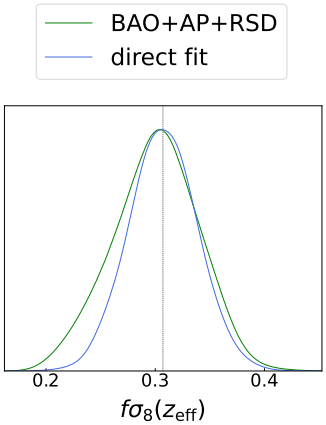

Figure 5.4.2: Posterior plot for $f\sigma_8(z_{\rm eff})$, comparing our method ('direct fit') in blue against C21 ('BAO+AP+RSD') in green.

Such an improvement was already found by C21, who demonstrated that considering the AP effect from the *smooth* component in addition to the *peak* provides significantly tighter constraints. We present results using their method with the additional RSD information, and we will refer to it as 'BAO+AP+RSD'. By running their analysis over our noiseless data, we find that 'BAO+AP+RSD' is able to place a constraint on $H_0 r_{\rm d}/c$ of 1.65%, which is of the same order as for our analysis (Tab. 5.4.1). Indeed, as it can be seen in Fig. 5.4.1, the green and the blue contours are both significantly tighter than the yellow one by almost the same amount and a consistent correlation coefficient of $\sim$ 0.95 between the two parameters is found. Despite both approaches being significantly more informative than the 'BAO' only analysis, a difference of about 16% and 17% is found in $H_0 r_{\rm d}$ and $\Omega_{\rm M}$ respectively. However, given that the ultimate goal is to demonstrate the difference between the 'direct fit' and the 'BAO' only approaches, investigating the less marked difference between the blue and green contours was outside the interest of this work. Our method constrains the growth of structure (Eq. 5.5-5.6) as well. For this reason, we also compare the RSD information of C21 with ours. The $f\sigma_8(z_{\rm eff})$ one-dimensional marginals for both methods are plotted in Fig. 5.4.2 and the constraining power is given for completion in Tab. 5.4.1. Overall, as already highlighted above, our method provides tighter constraints with respect to the

'BAO+AP+RSD' analysis. This is also the case for $f\sigma_8(z_{\rm eff})$, for which the constraint get tighter by about 18%.

5.4.2 Direct fit analysis without RSD

In order to further investigate where some of the information in the 'direct fit' alone is coming from, we repeat the same analysis, marginalizing over the growth of structure. We refer to this case as 'direct fit without RSD'. If we use the quasar RSD parameter defined as $\beta_{\rm QSO} = f(z)/b_{\rm QSO}$, Eq. (5.6) can be rewritten as

$$\begin{aligned} P_\times(k,\mu_k,z) =& b_{\rm Ly\alpha}\left(1+\beta_{\rm Ly\alpha}\mu_k^2\right) \\ & \times b_{\rm QSO}\left(1+\beta_{\rm QSO}\mu_k^2\right)F_{\rm nl,QSO}P(k,z)\ , \end{aligned} \tag{5.9}$$

where $\beta_{\rm QSO}$ is now an extra nuisance parameter.

The implications of this choice can be directly seen in Fig. 5.4.3. In the no-RSD case (red in Figure), the degeneracy among $A_{\rm s}$ and the two linear biases extends up to the prior limits, while the information on the other cosmological parameters is preserved. Marginalizing over the growth of structure washes out any constraining power on the linear biases and $A_{\rm s}$, while RSD are not constraining any other parameter.

5.5 Conclusions

Given that baryon acoustic oscillations (BAO) produce a distinct feature in 3D Lyα correlation functions, and its properties are well understood, BOSS and eBOSS analyses so far considered only the peak for cosmological inference ('BAO' analysis). A previous analysis conducted by Cuceu et al. (2021) (C21 throughout the paper) highlighted the importance of also considering the broadband component, as it significantly contributes to the overall cosmological information via the Alcock-Paczyński (AP) effect ('BAO+AP' analysis – 'BAO+AP+RSD' if redshift space

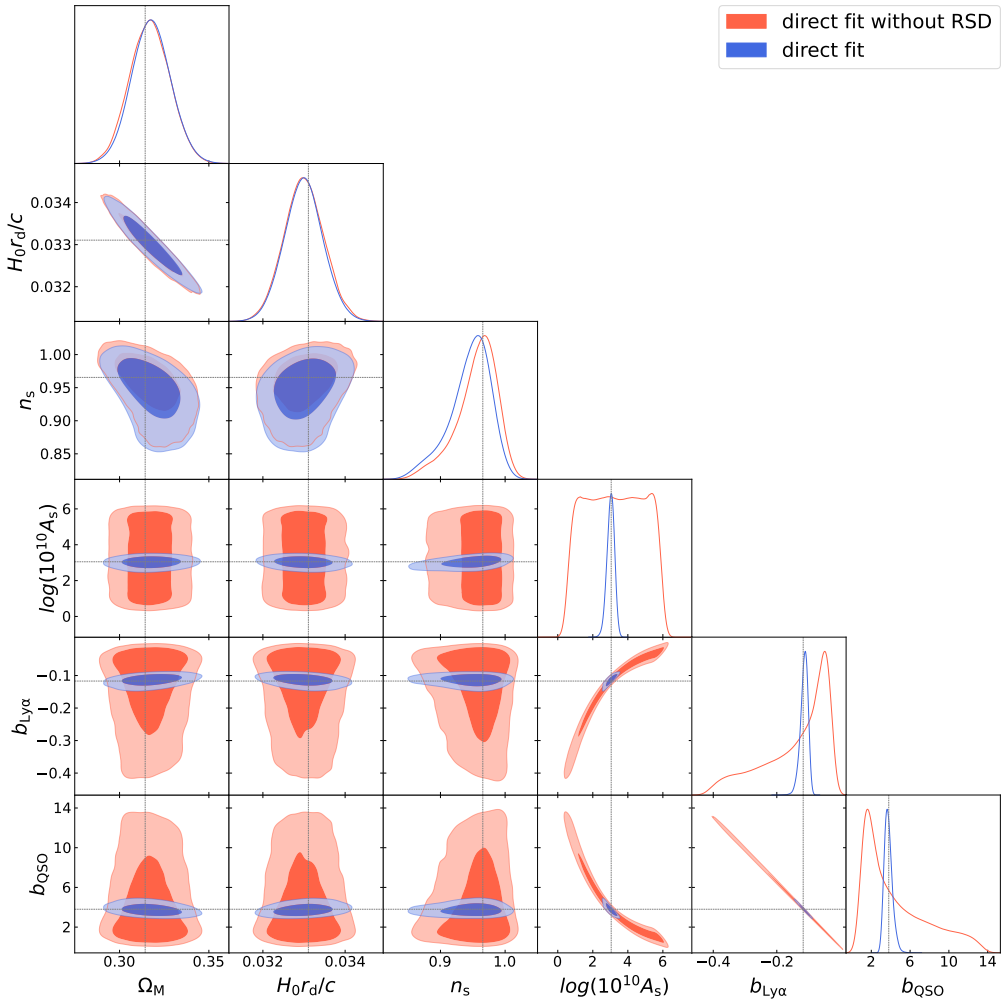

Figure 5.4.3: Triangle plot comparing the constraints on $\{\Omega_\mathrm{M}, H_0 r_\mathrm{d}/c, A_\mathrm{s}, n_\mathrm{s}\}$ and the two linear biases $\{b_{\mathrm{Ly}\alpha}, b_\mathrm{QSO}\}$ using our 'direct fit' approach in blue against the same fitting method but marginalizing over the growth of structure ('direct fit without RSD') in red. A further explanation of the second approach can be found in Sect. 5.4.2.

distortions (RSD) are included). Given these premises, in this paper we addressed the question about whether or not the compressed analyses based on BAO, AP and RSD parameters are able to capture all the cosmological information brought by the Lyα correlation functions. We performed a full shape analysis without any of the above parameters and instead directly inferred cosmology ('direct fit' analysis).

The inference framework we used is COBAYA, for which we implemented an *ad-hoc* gaussian likelihood based on the VEGA package, as extensively described in Sect. 5.2.

We performed 'BAO', 'BAO+AP+RSD' and 'direct fit' analyses on the same set of synthetic Ly$\alpha\times$Lyα auto- and Ly$\alpha\times$QSO cross-correlations, which include distortion effects due to continuum fitting, but no other contaminants, and ran the inference over $\boldsymbol{p_C} = \{H_0, \Omega_\mathrm{M}, \Omega_\mathrm{B}h^2, A_\mathrm{s}, n_\mathrm{s}\}$. We also marginalized over the nuisance astrophysical parameters $\boldsymbol{p_A} = \{b_{\mathrm{Ly}\alpha}, b_\mathrm{QSO}, \beta_{\mathrm{Ly}\alpha}, \sigma_\mathrm{v}\}$, which are the Ly$\alpha$ and quasars linear biases, the Lyα RSD term and the velocity dispersion of quasars respectively.

We were able to measure the matter density parameter Ω_M and the amplitude of primordial fluctuations A_s with a precision of 3.5% and 21%, respectively, and $f\sigma_8(z_\mathrm{eff})$ at 10.4%, which is a noteworthy result given the few measurements of this parameter at $z > 2$. For these parameters we do not obtain a significant improvement in constraining power with respect to the 'BAO+AP+RSD' approach. However, we obtain a constraint of the spectral index n_s at the $\sim 3-4\%$ level. Similarly to the findings and solutions put forward in Brieden et al. (2021), we could account for this extra information by adding a slope parameter to the compressed analysis.

The robustness of the 'direct fit' method against systematics must be tested. In forthcoming work, it would be interesting to investigate whether the compressed analysis is more robust, given it measures specific physical effects that are well understood. A further natural next step should be including contaminants to the analysis. The constraining power in the spectral index that we achieve could be affected in particular by Damped Lyα System (DLA) contamination (McQuinn and White, 2011; Font-Ribera et al., 2012) and fluctuations in the UV background (Pontzen and Governato, 2014; Gontcho et al., 2014), which would change the correlation function in a way

that could mimic $n_{\rm s}$. Further improvements to the analysis could come from varying $r_{\rm min}$ and $r_{\rm max}$, and also checking how effects of continuum distortion consequently behave.

We conclude by recalling that soon DESI will provide even better measurements of Lyα correlations. Therefore this kind of study is key to finding the optimal approach to infer cosmology from the data.

The code is publicly available at https://github.com/frgerardi/LyA_directfit.

6 Optimal data compression for Lyman-α forest cosmology

The Lyman-α (Lyα) three-dimensional correlation functions have been widely used to perform cosmological inference using the baryon acoustic oscillation (BAO) scale. While the traditional inference approach employs a data vector with several thousand data points, we apply near-maximal score compression down to tens of compressed data elements. We show that carefully constructed additional data beyond those linked to each inferred model parameter are required to preserve meaningful goodness-of-fit tests that guard against unknown systematics, and to avoid information loss due to non-linear parameter dependencies. We demonstrate, on suites of realistic mocks and DR16 data from the Extended Baryon Oscillation Spectroscopic Survey, that our compression approach is lossless and unbiased, yielding a posterior that is indistinguishable from that of the traditional analysis. As an early application, we investigate the impact of a covariance matrix estimated from a limited number of mocks, which is only well-conditioned in compressed space.

The work presented in this chapter has been published, see Gerardi et al. (2023).

6.1 Introduction

In recent decades, the Lyman-α (Lyα) forest gained popularity as a probe of the distribution of matter at redshifts $z > 2$. The forest consists of a sequence of absorption lines in high-redshift quasar (QSO) spectra, caused by neutral hydrogen placed along the line-of-sight, and hence it is a tracer of the intergalactic medium (IGM). Therefore, it contains cosmological information, and in particular Lyman-α clustering shows the distinct baryon acoustic oscillations (BAO) feature. This feature was first detected in the Lyα auto-correlation function using the Baryon Oscillation Spectroscopic Survey (BOSS) DR9 data (Busca et al., 2013; Slosar et al., 2013; Kirkby et al., 2013),

and subsequently extracted from the Lyα cross-correlation with QSOs using DR11 data (Font-Ribera et al., 2014).

The Lyα forest auto-correlation and its cross-correlation with quasars have been widely used to place constraints on the cosmological model (e.g., Aubourg et al., 2015; Alam et al., 2017; Cuceu et al., 2019; Alam et al., 2021; Cuceu et al., 2023). These two correlation functions are typically computed on a 2D grid in comoving coordinates along and across the line-of-sight, resulting in high dimensional data vectors, usually 2500 long for the auto-correlation and 5000 for the cross-correlation. However, standard BOSS and eBOSS (du Mas des Bourboux et al. 2020; hereafter dMdB20) Lyα forest analyses have so far focused on extracting cosmological information from the BAO peak, which is well localized to a smaller subset of bins. This means that the vector can be reduced to a smaller dimensionality, encoding the information we wish to capture. Hence, in this context, applying a data compression scheme could be useful to optimize the inference. In addition, the accuracy of the parameter estimates is tightly linked to the covariance matrix of the data vector, under the assumption of a Gaussian likelihood. As the true covariance $\mathbf{\Sigma}$ of the correlation function is inaccessible, standard analyses usually estimate it either from large set of mocks or analytically from models of the covariance matrix (Kitaura et al., 2016; Wadekar et al., 2020). In Lyα analyses, producing mocks can be a highly computationally-expensive process, therefore only a limited number is available, 100 in the case of dMdB20. However, if the number of samples is significantly lower than the number of data points, the estimate of the covariance is singular and has no inverse (Hartlap et al., 2007; Dodelson and Schneider, 2013; Taylor and Joachimi, 2014; Sellentin and Heavens, 2015; Percival et al., 2021).

In the eBOSS DR16 analysis, the covariance matrix $\mathbf{C}$ is computed via the sub-sampling method, which, given some dataset, consists of computing the covariance of correlation functions obtained in individual subsamples of the sky. Despite being larger ($\sim$ 800) than the number of mocks (100), the number of subsamples is still lower than the number of data points (2500-5000); hence, the covariance matrix must be tested. Alternatively, in the same analysis, the authors computed a Gaussian covariance matrix using the Wick approximation (Delubac et al., 2015) and used it to

benchmark the covariance computed from the sub-sampling method. The accuracy of the covariance matrix would increase by alleviating the mismatch between the number of bins and the number of mocks. This can be done by applying a data compression algorithm and evaluating the (compressed) data covariance matrix in a new space characterized by a lower dimensionality. In particular, given the available set of a hundred mocks, we reduce each of them to a set of compressed data vectors and compute a newly defined mock sample covariance, which is a good estimator of the true covariance, given that the length of the compressed data vector is now much smaller than the number of mocks. Then, a comparison between the covariance matrix of the data, mapped into the compressed space, and the mock sample covariance, obtained from the compressed vector, can clarify whether there has been an underestimation or overestimation of the contours in the standard analyses. Moreover, we are interested in obtaining a more sensitive goodness of fit test. The length of Lyα correlation data vectors is of the order of $\mathcal{O}(10^3)$, which could easily hide any bad fit in a subset of the data. By reducing the dimensionality of the data vector through compression, we wish to obtain a test that would highlight when a few important points are off.

Driven by these optimization problems, we perform the inference analysis on realistic Ly$\alpha\times$Lyα auto- and Ly$\alpha\times$QSO cross-correlation functions in a data compression framework. The compression algorithm we use is *score compression* (Alsing and Wandelt, 2018), under the hypothesis of a Gaussian likelihood (and hence analogous to the MOPED scheme; see Heavens et al. 2000). By construction, the dimensionality of the compressed data vector will be equal to the number of parameters we wish to keep information of, namely $\mathcal{O}(10)$.

The paper is structured as follows. We start in Sect. 6.2 by outlining the method, explaining the computation of the covariance matrix, and introducing the modelling and the basic idea behind score compression. We proceed in Sect. 6.3 by testing the compression algorithm against loss of information, comparing the inferred posterior distribution for our sampled parameters in the traditional and compressed frameworks. In Sect. 6.4, we compare the constraining power of the original estimated covariance matrix against the mock-to-mock covariance. We then per-

form goodness of fit tests in the compressed framework in Sect. 6.5. Throughout the analysis a tight prior on the BAO parameters is imposed to overcome the problem of the non-linear relation between these and their corresponding summary statistics components. We relax the prior constraint, and hence made the analysis more generalizable, by extending the framework as described in Sect. 6.6. An application of our new framework to eBOSS DR16 data is presented in Sect. 6.7. Conclusions are drawn in Sect. 6.8.

Making sure the analysis is both optimized and reliable is key to interpret the vast amount of Lyα forest data which will become available from the Dark Energy Spectroscopic Instrument (DESI).

6.2 Method

Generically referring to the Lyα auto- and cross-correlations as the data vectors, the goal of this work is to study data compression in the context of Lyα forest 3D analyses. In particular, this means compressing the data down to a set of summary statistics $\mathbf{t}$, which will encode into a shorter vector the information we are interested in. As we have just seen, this also benefits the computation of the covariance matrix. The new 'compressed' framework is tested against the traditional analysis while performing parameter inference. To evaluate posterior distributions we use the nested sampler Polychord (Handley et al., 2015a,b).

We start in Sect. 6.2.1 by introducing the mocks used in this analysis, with a focus on the computation of the covariance matrix. We then describe the modelling of the Ly$\alpha\times$Lyα and the cross Ly$\alpha\times$QSO power spectra in Sect. 6.2.2, as implemented in vega[1] (Cuceu et al., 2023), and the set of randomly generated *Monte Carlo realizations* of the correlation function in Sect. 6.2.3. In Sect. 6.2.4 we finally outline the compression method used, namely *score compression*.

[1] https://github.com/andreicuceu/vega

6.2.1 Synthetic data vector and covariance

In this work we use a set of 100 realistic Lyα mocks, with and without contaminants, which were produced for the Lyα eBOSS DR16 analysis (du Mas des Bourboux et al., 2020). The synthetic Lyα transmitted fluxes are produced using the CoLoRe (Ramírez-Pérez et al., 2022) and LyaCoLoRe (Farr et al., 2020) packages, from the same cosmology for all the mocks. Synthetic quasar spectra are then generated given some astrophysical and instrumental prescriptions, and contaminants are added if requested. Then the mocks run through the same analysis pipeline (picca[2]) as the real data, resulting in measured auto- and cross-correlation functions (dMdB20). These are derived from computing the correlation function in each HEALPix[3] (Górski et al., 2005) pixel — about 880 pixels (subsamples) for the eBOSS footprint (NSIDE=16) — and evaluating the mean and covariance over the full set of pixels of the mock, to be then assigned to the entire survey. In this way, for every i-th mock, there will be a measurement of both the correlation function and the covariance matrix $\boldsymbol{C}_i$, which will be only an estimate of the true covariance $\boldsymbol{\Sigma}$ as mentioned above. In each subsample, the correlation has a size of either 2500 ($\xi_{\rm auto}$) or 5000 ($\xi_{\rm cross}$) bins, hence the number of subsamples (880 pixels) is significantly lower than the number of data points (2500 or 5000). This means that the covariance should be singular, however off-diagonal elements of the correlation matrix are smoothed to make it positive definite (dMdB20).

Finally, given the same hundred mocks, it is possible to define a *stack* of them. In particular, the correlation function for the *stack* of mocks is obtained by collecting all the subsamples (for all the hundred mocks), and computing the mean and covariance of the correlation functions computed in each of them, effectively reducing the noise. We will refer to the contaminated auto- and cross-mock correlations of the *stack* as *stacked correlations*.

In this analysis, we use the same scale cuts as in eBOSS DR16 (du Mas des Bourboux et al., 2020), assuming $r_{\rm min} = 10{\rm h}^{-1}{\rm Mpc}$, up to $r_{\rm max} = 180{\rm h}^{-1}{\rm Mpc}$. The effective redshift of the correlation functions is $z_{\rm eff} = 2.3$.

[2] https://github.com/igmhub/picca
[3] https://healpix.sourceforge.io

			Testing the framework (*stacked*)		Testing the covariance (single mock)	
Parameter	Fiducial	Prior	Traditional	Compression	Original cov.	Mock-to-mock cov.
$\alpha_\parallel$	1.00	$\mathcal{U}(0.88, 1.14)$	1.000 ± 0.002	1.000 ± 0.002	1.003 ± 0.019	1.003 ± 0.019
$\alpha_\perp$	1.01	$\mathcal{U}(0.88, 1.14)$	1.004 ± 0.003	1.004 ± 0.003	1.002 ± 0.027	$1.004^{+0.029}_{-0.032}$
$b_{\mathrm{Ly}\alpha}$	-0.14	$\mathcal{U}(-2, 0)$	-0.135 ± 0.001	-0.135 ± 0.001	-0.125 ± 0.004	-0.124 ± 0.006
$\beta_{\mathrm{Ly}\alpha}$	1.41	$\mathcal{U}(0, 5)$	1.47 ± 0.01	1.47 ± 0.01	$1.67^{+0.07}_{-0.08}$	$1.68^{+0.09}_{-0.10}$
b_{QSO}	3.81	$\mathcal{U}(0, 10)$	3.80 ± 0.01	3.80 ± 0.01	3.82 ± 0.08	3.81 ± 0.07
β_{QSO}	0.25	$\mathcal{U}(0, 5)$	0.25 ± 0.01	0.25 ± 0.01	0.27 ± 0.04	$0.27^{+0.03}_{-0.04}$
$\sigma_{\mathrm{v}}(\mathrm{M}pc/h)$	2.87	$\mathcal{U}(0, 15)$	2.82 ± 0.04	2.82 ± 0.04	$3.22^{+0.32}_{-0.28}$	3.24 ± 0.26
$\sigma_{\parallel,\mathrm{sm}}$	2.05	$\mathcal{U}(0, 10)$	2.08 ± 0.01	2.08 ± 0.01	2.10 ± 0.09	$2.10^{+0.09}_{-0.08}$
$\sigma_{\perp,\mathrm{sm}}$	2.35	$\mathcal{U}(0, 10)$	2.33 ± 0.01	2.33 ± 0.01	2.23 ± 0.11	2.21 ± 0.11
$b_{\mathrm{HCD}}[\times 10^{-2}]$	-1.70	$\mathcal{U}(-20, 0)$	-2.12 ± 0.08	-2.13 ± 0.07	-2.98 ± 0.54	-3.06 ± 0.68
β_{HCD}	1.57	$\mathcal{N}(0.5, 0.09)$	0.86 ± 0.06	0.86 ± 0.06	0.50 ± 0.09	0.50 ± 0.09
$b_{\eta,\mathrm{SiII(1260)}}[\times 10^{-3}]$	-0.58	$\mathcal{U}(-50, 50)$	-0.59 ± 0.04	-0.59 ± 0.04	-0.83 ± 0.33	-0.88 ± 0.37
$b_{\eta,\mathrm{SiII(1193)}}[\times 10^{-3}]$	-1.12	$\mathcal{U}(-50, 50)$	-1.09 ± 0.03	-1.09 ± 0.03	-0.83 ± 0.27	-0.84 ± 0.28
$b_{\eta,\mathrm{SiIII(1207)}}[\times 10^{-3}]$	-1.75	$\mathcal{U}(-50, 50)$	-1.64 ± 0.03	-1.63 ± 0.03	-1.54 ± 0.31	-1.52 ± 0.30
$b_{\eta,\mathrm{SiII(1190)}}[\times 10^{-3}]$	-1.00	$\mathcal{U}(-50, 50)$	-1.00 ± 0.03	-1.00 ± 0.03	-0.75 ± 0.27	-0.75 ± 0.29

Table 6.2.1: Full set of sampled parameters, alongside with the fiducial values used to compute the summary statistics (see Eq. 6.8), priors and the 1-D marginals (68% c.l.). Uniform ($\mathcal{U}$) priors adopted for the sampling procedure, while we assign a Gaussian prior on β_{HCD}, where by notation the Gaussian distribution $\mathcal{N}(\mu, \sigma)$ has mean μ and standard deviation σ. Results are split into 'Testing the framework (*stacked*)' and 'Testing the covariance (single mock)', which respectively refer to the setup in Sect. 6.3 and Sect. 6.4. The former set of results shows the comparison between the traditional and the compressed inference pipelines using the *stacked* auto- and cross-correlation mocks, while the second between the compressed method using either the original covariance $\boldsymbol{C}$ (which is mapped into the compressed space) or the mock-to-mock covariance $\boldsymbol{C_t}$, for a single mock.

6.2.2 Modelling and parameter space

To model the Lyα correlation functions we follow Eq. (27) of du Mas des Bourboux et al. (2020), while applying the same prescriptions as in Gerardi et al. (2022) (see Chapter 5). Given a certain cosmological model and a corresponding isotropic linear matter power spectrum $P(k, z)$, the Lyα auto and Lyα-QSO cross power spectra are computed as

$$P_{\mathrm{Ly}\alpha}(k, \mu_k, z) = b_{\mathrm{Ly}\alpha}^2 \left(1 + \beta_{\mathrm{Ly}\alpha}\mu_k^2\right)^2 F_{\mathrm{nl,Ly}\alpha}^2(k, \mu_k) P(k, z) \; ; \tag{6.1}$$

$$\begin{aligned} P_{\times}(k, \mu_k, z) =& b_{\mathrm{Ly}\alpha}\left(1 + \beta_{\mathrm{Ly}\alpha}\mu_k^2\right) \\ & \times b_{\mathrm{QSO}}\left(1 + \beta_{\mathrm{QSO}}\mu_k^2\right) F_{\mathrm{nl,QSO}}(k_\parallel) P(k, z) \; , \end{aligned} \tag{6.2}$$

where $\mu_k = k_\parallel / k$, with k and $k_\parallel$ the wave vector modulus and its line-of-sight component, respectively. On one hand, the Ly$\alpha\times$Lyα power spectrum in Eq. (6.1) depends on the Lyα forest linear bias $b_{\mathrm{Ly}\alpha}$ and RSD parameter $\beta_{\mathrm{Ly}\alpha} = \frac{b_{\eta,\mathrm{Ly}\alpha} f(z)}{b_{\mathrm{Ly}\alpha}}$, where $b_{\eta,\mathrm{Ly}\alpha}$ is an extra unknown bias, the velocity divergence bias, and $f(z)$ the logarithmic growth rate. The $F_{\mathrm{nl,Ly}\alpha}$ term accounts for non-linear corrections (Arinyo-i-Prats et al., 2015). On the other hand, the quasar parameters that contribute to the Ly$\alpha\times$QSO power spectrum in Eq. (6.2) are the quasar linear bias b_{QSO} and the redshift-space distortions (RSD) term $\beta_{\mathrm{QSO}} = f(z)/b_{\mathrm{QSO}}$. Finally, we model non-linear effects of quasars and redshift errors following du Mas des Bourboux et al. (2020), using a Lorentzian function

$$F_{\mathrm{nl,QSO}}(k_\parallel) = \left[1 + \left(k_\parallel \sigma_{\mathrm{v}}\right)^2\right]^{-1/2} , \tag{6.3}$$

where σ_{v} is the velocity dispersion.

The power spectra in Eqs. (6.1-6.2) only account for Lyα flux and in reality this is also contaminated by absorption lines due to heavy elements, generally referred to as metals, and high column density (HCD) systems (Bautista et al., 2017; Font-Ribera et al., 2012). Let us first focus

on the modelling of the HCDs. Font-Ribera et al. (2012) showed their broadening effect along the line-of-sight can be modeled at the level of new effective Lyα bias and RSD parameters

$$b'_{\mathrm{Ly}\alpha} = b_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}} F_{\mathrm{HCD}}(k_{\parallel}) \,, \tag{6.4}$$

$$b'_{\mathrm{Ly}\alpha}\beta'_{\mathrm{Ly}\alpha} = b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}}\beta_{\mathrm{HCD}} F_{\mathrm{HCD}}(k_{\parallel}) \,, \tag{6.5}$$

with b_{HCD} and β_{HCD} being the linear bias and RSD parameters. $F_{\mathrm{HCD}}(k_{\parallel})$ is a function of the line-of-sight wavenumber, and it is modeled following dMdB20. On the other hand, metals contribute to the final auto- and cross-correlation functions as per

$$\xi'_{\mathrm{auto}} = \xi_{\mathrm{Ly}\alpha\times\mathrm{Ly}\alpha} + \sum_{m} \xi_{\mathrm{Ly}\alpha\times m} + \sum_{m_1,m_2} \xi_{m_1\times m_2} \,, \tag{6.6}$$

$$\xi'_{\mathrm{cross}} = \xi_{\mathrm{Ly}\alpha\times\mathrm{QSO}} + \sum_{m} \xi_{\mathrm{QSO}\times m} \,, \tag{6.7}$$

where m generically refer to a metal and the sums are performed over all possible metals considered. The modelling of the cross-correlation of a metal with other metals ($\xi_{m_1\times m_2}$) and with Lyα ($\xi_{\mathrm{Ly}\alpha\times m}$) and QSO ($\xi_{\mathrm{QSO}\times m}$) follows the modelling of the auto- and cross-correlations of the Lyα, and each metal line has a linear bias b_m and RSD parameter $\beta_m = b_{\eta,m} f(z)/b_m$. Following dMdB20, we fix all $\beta_m = 0.5$, and sample the metal biases.

Based on this modelling, we use the code VEGA to compute the two-dimensional correlation functions $\boldsymbol{\xi}$. This same code computes both the BAO feature parameters $\{\alpha_{\parallel}, \alpha_{\perp}\}$, which shift the peak along and across the line-of-sight, and the Gaussian smoothing (Farr et al., 2020), which accounts for the low resolution of the mocks and is parameterized by $\{\sigma_{\parallel}, \sigma_{\perp}\}$ smoothing parameters.

At the inference level, the set of sampled parameters is $\boldsymbol{p}_{\mathbf{s}} = \{\alpha_{\parallel}, \alpha_{\perp}, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}, \sigma_{\parallel}, \sigma_{\perp}\}$, which is extended to include also $\{b_{\eta,\mathrm{m}}, b_{\mathrm{HCD}}, \beta_{\mathrm{HCD}}\}$ when also fitting for contaminants. In this notation, $b_{\eta,\mathrm{m}}$ is the velocity divergence bias for the metal m — here we consider SiII(1260), SiII(1193), SiIII(1207) and SiII(1190).

For all these parameters we choose uniform priors, which are listed in Tab. 6.2.1. The only exception is given by $\beta_{\rm HCD}$, for which, following the previous eBOSS DR16 analysis, we impose an informative Gaussian prior.

6.2.3 Monte Carlo realizations

We here introduce a different kind of simulated data, which we will later use, defined as *Monte Carlo realizations*. They are correlation functions obtained by adding noise on top of the model, as defined in Sect. 6.2.2. The noise is given by a covariance matrix from one of the hundred mocks correlation that have been seen so far. What this means is that we can imagine every data point to be generated from a normal distribution $\mathcal{N}(\boldsymbol{\xi}, \boldsymbol{C})$, where $\boldsymbol{\xi}$ is the model correlation function and $\boldsymbol{C}$ is given by the covariance of the first realistic mock. Using Monte Carlo simulations comes with two advantages. First, it is possible to generate as many as needed to build any statistics. Secondly, we have control over the model and there will be clear knowledge of the underlying physics.

6.2.4 Score compression

To reduce the dimensionality of our datasets we use score compression (Alsing and Wandelt, 2018). Given a known form for the log-likelihood function $\mathcal{L}$, this method corresponds to linear transformations of the data, based on the idea of compressing them down to the score function $\boldsymbol{s} = \nabla \boldsymbol{\mathcal{L}}_*$. The components of the compressed vector are the derivatives of the log-likelihood function, evaluated at some fiducial set of parameters $\boldsymbol{\theta}_*$, with respect to the parameters of interest $\boldsymbol{\theta}$. Under the assumptions that the likelihood function is Gaussian and the covariance $\boldsymbol{C}$ does not depend on parameters, from the data $\boldsymbol{d}$ the compressed data vector is obtained as

$$\boldsymbol{t} = \nabla \boldsymbol{\mu}_*^T \boldsymbol{C}^{-1} (\boldsymbol{d} - \boldsymbol{\mu}_*) \,, \tag{6.8}$$

where $\boldsymbol{\mu}_*$ is the fiducial model. Under these assumptions the compression is identical to the widely used MOPED scheme (Heavens et al., 2000) apart from a bijective linear transformation.

In our case the model corresponds to the correlation function $\boldsymbol{\xi}$, described earlier in Sect. 6.2.2. The corresponding likelihood distribution in compressed space will be then given by

$$P(\boldsymbol{t}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{F}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}[\boldsymbol{t}-\boldsymbol{\mu_t}(\boldsymbol{\theta})]^T \boldsymbol{F}^{-1}[\boldsymbol{t}-\boldsymbol{\mu_t}(\boldsymbol{\theta})]\right] , \tag{6.9}$$

where n is the length of $\boldsymbol{t}$, $\boldsymbol{\mu_t}(\boldsymbol{\theta})$ is the compressed model $\boldsymbol{\mu}$ evaluated at $\boldsymbol{\theta}$, namely $\boldsymbol{\mu_t}(\boldsymbol{\theta}) = \nabla\boldsymbol{\mu}_*^T \boldsymbol{C}^{-1}[\boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\mu}_*]$, and

$$\boldsymbol{F} = [\nabla\boldsymbol{\mu}_*]^T \boldsymbol{C}^{-1}[\nabla^T\boldsymbol{\mu}_*] \tag{6.10}$$

is the Fisher matrix.

When considering both the auto- and cross-correlations, some parameters will be in common; for this reason, there is the need to build a joint summary statistic. If we define independently the Lyα auto- and cross- data vectors, characterized by the covariances $\mathbf{C}_{\rm auto}$ and $\mathbf{C}_{\rm cross}$ respectively, and given they do not correlate with each other, in the joint analysis the full covariance matrix will be given by

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{\rm auto} & 0 \\ 0 & \mathbf{C}_{\rm cross} \end{pmatrix} . \tag{6.11}$$

Then the resulting summary statistics vector and Fisher matrix will be respectively obtained as $\boldsymbol{t} = \boldsymbol{t}_{\rm auto} + \boldsymbol{t}_{\rm cross}$ and $\boldsymbol{F} = \boldsymbol{F}_{\rm auto} + \boldsymbol{F}_{\rm cross}$.

This compression method is dependent on the choice of the fiducial set of parameters $\boldsymbol{\theta}_*$, which might not be known *a priori*. However, Alsing and Wandelt (2018) suggest iterating over the *Fisher scoring method* for maximum-likelihood estimation

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \boldsymbol{F}_k^{-1}\nabla\boldsymbol{\mathcal{L}}_k , \tag{6.12}$$

until convergence of the full set of parameters. How this is done in our particular case is described at the beginning of Sect. 6.3. An important note is that this iterative procedure does not take

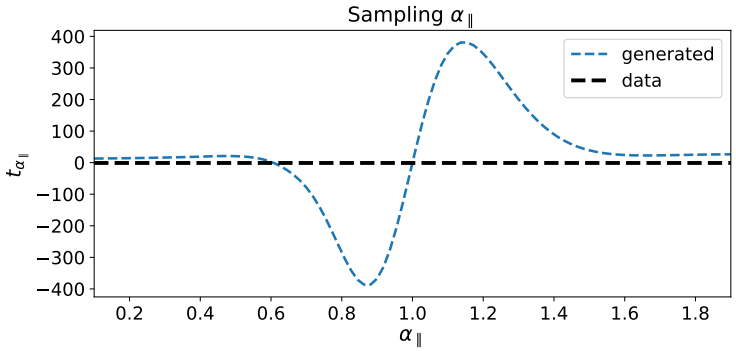

Figure 6.2.1: This plot shows the behaviour of the summary component $t_{\alpha_\parallel}$ as a function of $\alpha_\parallel$, which is the parameter it is related to as per Eq. (6.8), against the value of $t_{\alpha_\parallel}$ evaluated using $\alpha_\parallel = 1.00$ (see Tab. 6.2.1), denoted as 'data'. The remainder of the parameters are set to the fiducial values listed in Tab 6.2.1. This figure highlights a non-monotonic relationship between the two parameters, which would lead to multiple peaks in the posterior if a tight prior is not imposed.

into account parameters priors, which means it can potentially lead to unusual values for those parameters which are not well constrained by the data.

In computing the score compression components over the parameters $\{\alpha_\parallel, \alpha_\perp\}$, we realized their relation with their corresponding summary statistics components, namely $\{\boldsymbol{t}_{\alpha_\parallel}, \boldsymbol{t}_{\alpha_\perp}\}$, was not monotonic, as per Fig. 6.2.1. This is problematic as this means the posterior can have more than one peak (Graff et al., 2011) if we sample over the full $[0.01, 1.99]$ interval. We overcame this complexity by imposing a tighter prior on $\{\alpha_\parallel, \alpha_\perp\}$ at the sampling step. This prior is designed to allow for $\alpha_\parallel$ values in between the minimum and maximum of $\boldsymbol{t}_{\alpha_\parallel}(\alpha_\parallel)$. The same prior is imposed on $\alpha_\perp$. This tightening does not affect the inference when performed on the correlation function of the *stacked* mock, in which case posteriors are well within this prior, but it reveals to be quite important when evaluating the posteriors on the individual mocks. For this reason, we make sure we provide example results for those mocks whose contours are within the prior range.

Later, in Sect. 6.6 we will see how we can remove the tight prior constraint by evaluating the summary statistics components associated to $\{\alpha_\parallel, \alpha_\perp\}$ at multiple fiducial values of the BAO parameters, effectively enlarging the compressed vector.

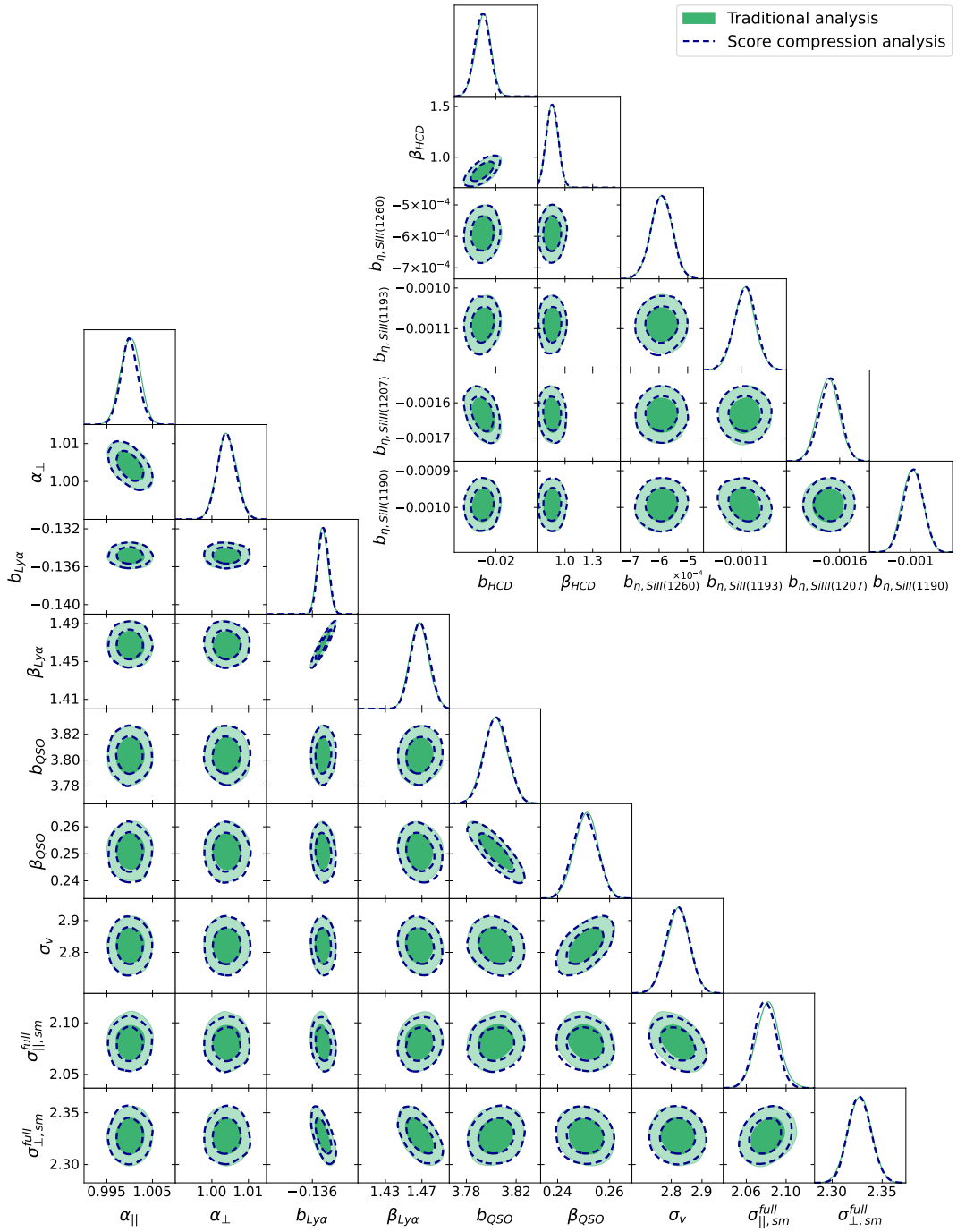

Figure 6.2.2: Triangle plots of the parameters of interest for the *stack* of correlation functions computed from a set of 100 mocks. Results are split, for presentation purposes only, into the set of standard parameters $\{\alpha_{\parallel}, \alpha_{\perp}, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}, \sigma_{\parallel}, \sigma_{\perp}\}$ (lower left panel) and contaminants parameters $\{b_{\eta,\mathrm{SiII}(1260)}, b_{\eta,\mathrm{SiII}(1193)}, b_{\eta,\mathrm{SiIII}(1207)}, b_{\eta,\mathrm{SiII}(1190)}, b_{\mathrm{HCD}}, \beta_{\mathrm{HCD}}\}$ (upper right panel). The green contours refer to the results obtained performing the inference using the full uncompressed data vector, which we denote as 'Traditional analysis', while the blue dashed refer to the compressed analysis results, denoted as 'Score compression analysis'.

6.3 Compression performance

In this Section we apply the score compression algorithm, outlined in Sect 6.2.4, to Lyα auto- and cross-correlations measured from contaminated mocks. The pipeline starts by choosing a fiducial set of parameters for computing the score compressed vector, as per Eq. (6.8). The fiducial is obtained by iterating over Eq. (6.12), with $\boldsymbol{\theta}_0$ given by the best fit of the *stacked* correlation functions. Given this initial guess, we then iterated assigning to $\boldsymbol{\theta}_{k+1}$ the median of the $\boldsymbol{\theta}$ values over the hundred mocks at the k-th step.

The likelihood is assumed to be Gaussian, which has a major impact on the final form of the compressed vector, given that the latter is computed as the gradient of the log-likelihood. Based on previous analyses, we assume the data are normally distributed and this assumption of Gaussianity will also be inherited in the compressed space. In general, when mapping in a compressed space, this property might not be preserved, but given that score compression is a linear transformation, that is the case. We make a consistency check by running the Henze-Zirkler test (Henze and Zirkler, 1990) for multivariate normality in the compressed space. Intuitively, this test measures the distance between the measured and target (multivariate) distribution, and it was shown to perform well in high-dimensional problems. We found that the summary statistics, computed for the hundred mocks at the end of the iterative process, follows a multivariate normal distribution.

Provided the fiducial model and the Gaussianity checks, we first test the compression method on the *stack* of the mocks, with results presented in this Section, and later, in Sect. 6.4, we compute the covariance matrix for the summary statistics over the set of hundred mocks and compare it to the Fisher matrix as defined in Eq. (6.10). It is important to keep in mind that, when referring to the Fisher matrix, we are simply referring to the mapping of the data covariance matrix $\boldsymbol{C}$ into the compressed space.

To test the score compression algorithm against the traditional approach, for simplicity, we employ both the contaminated auto- and cross- *stacked correlations*, which are almost noise-free. This choice is motivated by the fact that we imposed a tight prior on the $\{\alpha_{\parallel}, \alpha_{\perp}\}$ parameters to

overcome the challenges coming from the non-monotonic relationship between these parameters and their corresponding summary statistics components (see Fig. 6.2.1). Thus, experimenting over less noisy mock data facilitates running the test in a case where it is granted that posteriors will not hit the priors.

For both the traditional (uncompressed data) and the compressed frameworks we run the PolyChord sampler for the auto- and cross-*stacked correlations*, while sampling the full set of 15 model parameters $\{\alpha_\parallel, \alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}, \sigma_\parallel, \sigma_\perp, b_{\eta,\mathrm{SiII(1260)}}, b_{\eta,\mathrm{SiII(1193)}}, b_{\eta,\mathrm{SiIII(1207)}}, b_{\eta,\mathrm{SiII(1190)}}, b_{\mathrm{HCD}}, \beta_{\mathrm{HCD}}\}$ and results are presented in Fig. 6.2.2. The two methods agree well with each other, leading to almost identical results. The numerical values of the peaks and 1σ confidence intervals of the 1d marginals are presented in Tab. 6.2.1 as part of the 'Testing the framework (*stacked*)' set of columns. From the table, it can be noticed that in some cases the fiducial parameters used to compute the compression are not within the 3σ confidence interval. Despite the fiducial being a first guess, and not necessarily accurate, the contours of the two methods agree well with each other.

We just demonstrated that the score compression inference pipeline leads to the same results as the standard approach. This shows the linearity of the parameters in the model to a good approximation. However, it is important to bear in mind that, in this case, this only holds locally around the fiducial, because of the non-linearity of the components that relate to $\alpha_\parallel$ and $\alpha_\perp$, on which we imposed a tight prior.

6.4 Testing the covariance matrix

An interesting application of the compression algorithm consists of evaluating the accuracy of the covariance matrix $\boldsymbol{C}$ by comparing it to the mock-to-mock covariance $\boldsymbol{C_t}$, which is the covariance matrix of the summary statistics vectors of the set of hundred mocks. We now showcase this application using a single mock.

We recall that the computation of the standard data covariance happens in a setup where the length of the data vector is larger than the number of samples, which is sub-optimal. The covari-

ance should be singular; however, the off-diagonal elements of the correlation matrix are smoothed to make it positive definite (du Mas des Bourboux et al., 2020). Reducing the dimensionality of the data vector via score compression allows us to compute a new covariance matrix $\boldsymbol{C_t}$, which has a dimensionality significantly lower than the number of samples used to compute it, given that the new data vector will be $\sim \mathcal{O}(10)$ long. The fact that now the number of mock samples is larger than the number of compressed data points, means that we are now in a framework where the estimated $\boldsymbol{C_t}$ is in principle a better estimator of the true covariance $\boldsymbol{\Sigma}$ in compressed space than $\boldsymbol{F}$, which is obtained by mapping the covariance $\boldsymbol{C}$ into this space.

We now repeat the same experiment as in Sect. 6.3 over a single mock and evaluate the posterior using PolyChord for the full set of parameters $\{\alpha_\parallel, \alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}, \sigma_\parallel, \sigma_\perp, b_{\eta,\mathrm{SiII}(1260)}, b_{\eta,\mathrm{SiII}(1193)}, b_{\eta,\mathrm{SiII}(1207)}, b_{\eta,\mathrm{SiII}(1190)}, b_{\mathrm{HCD}}, \beta_{\mathrm{HCD}}\}$. This is either done using the original covariance $\boldsymbol{C}$ matrix (mapped into the compressed space, to the Fisher matrix) in the Gaussian likelihood in Eq. (6.9) or instead using the mock-to-mock covariance $\boldsymbol{C_t}$ adopting a t-distribution as a likelihood function, as proposed in Sellentin and Heavens (2015). The latter is of the form of

$$P(\boldsymbol{t}|\boldsymbol{\theta}) = \frac{\bar{c}_P |\boldsymbol{C_t}|^{-1/2}}{1 + \frac{[\boldsymbol{t}-\boldsymbol{\mu_t}(\boldsymbol{\theta})]^T \boldsymbol{C_t}^{-1} [\boldsymbol{t}-\boldsymbol{\mu_t}(\boldsymbol{\theta})]}{n_s - 1}} \tag{6.13}$$

with

$$\bar{c}_P = \frac{\Gamma\left(\frac{n_s}{2}\right)}{[\pi(n_s-1)]^{n_t/2}\Gamma\left(\frac{n_s-n_t}{2}\right)} \; ; \tag{6.14}$$

where n_s is the number of mocks, n_t is the length of the compressed data vector and Γ is the Gamma function. Once again the choice of the tight prior on both $\{\alpha_\parallel, \alpha_\perp\}$ affected the choice of the set of mocks in order to run this second experiment. However, the goal of this second experiment is to provide an example case of testing the accuracy of the subsampling estimation of the covariance matrix. If the method is demonstrated to effectively work over some subset of mocks, it is expected that will also be the case for the others.

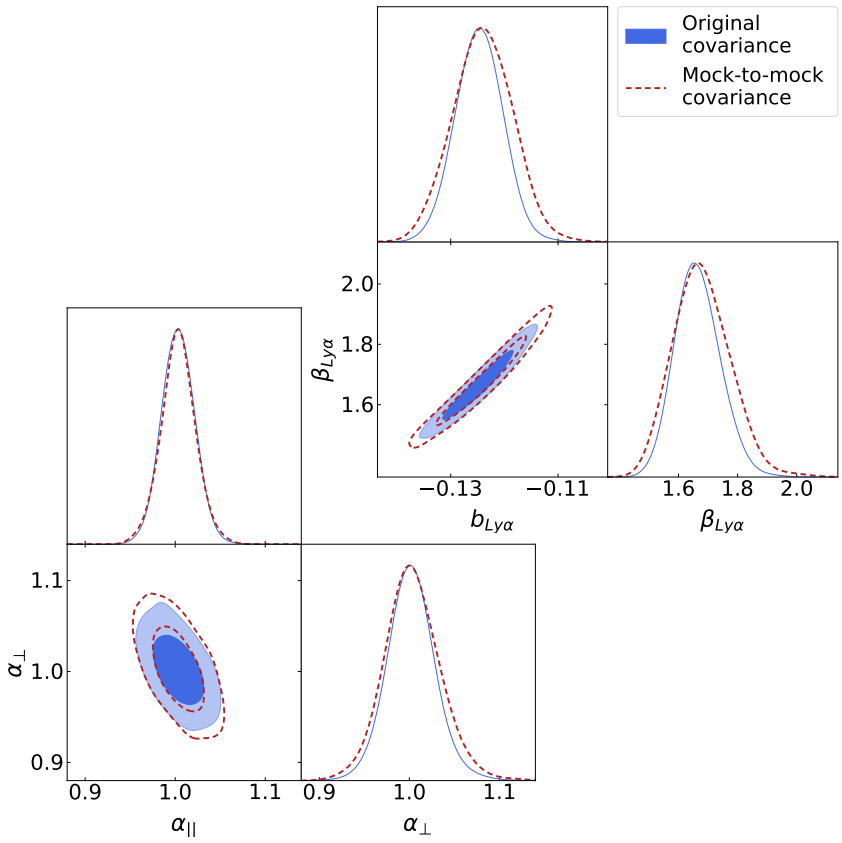

Figure 6.4.1: Triangle plots of the BAO parameters of interest $\{\alpha_\parallel, \alpha_\perp\}$ and the Lyα parameters $\{b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}\}$ for one set of the Lyα auto- and cross- mock correlations. The blue filled contours refer to the results obtained performing the inference using the original covariance matrix $\boldsymbol{C}$ (mapped into the compressed space) in the likelihood function, and hence are denoted as 'Original covariance'. On the other hand, the red dashed results, denoted as 'Mock-to-mock covariance', refer to the case in which the mock-to-mock covariance matrix is used instead, while adopting a t-distribution likelihood.

The results for the BAO parameters $\{\alpha_{\parallel}, \alpha_{\perp}\}$ and the Lyα parameters $\{b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}\}$ are shown in Fig. 6.4.1, while the full set is presented in Sect. 6.9 and listed in Tab. 6.2.1 ('Testing the covariance (single mock)' columns). While the blue contours refer to the results obtained using the original covariance matrix $\mathbf{C}$, the dashed red ones are achieved adopting a t-distribution likelihood combined with the mock-to-mock covariance $\mathbf{C}_t$. In this test case, using the mock-to-mock covariance results in a small enlargement of the posterior for the $\alpha_{\perp}$ parameter: while using the original covariance matrix provides $\alpha_{\perp} = 1.002 \pm 0.027$, the mock-to-mock covariance results in $\alpha_{\perp} = 1.004^{0.029}_{-0.032}$. On the other hand, the Lyα linear bias and RSD parameter absolute errors increase by 50% and $\sim 25\%$ respectively, with final relative error of about $5 - 6\%$. The uncertainty of the vast majority of the other parameters agree remarkably well.

We end this discussion on covariance matrix estimation by noting that the test presented here is meant as a showcase of the usefulness of compressing Lyα forest correlation functions. However, proper testing of the Lyα forest covariance matrices would require a more comprehensive analysis using a larger sample of mocks[4], and comparison with other estimation methods (see e.g., du Mas des Bourboux et al., 2020).

6.5 Goodness of fit test

In this section, we make a step forward with respect to the original aim of the work, by considering goodness of fit tests. For Lyα correlation functions, the length of the data vector can go from 2500, considering only the auto-, to 7500 if considering also the cross-correlation. In a context where only $\sim \mathcal{O}(10)$ parameters are sampled, any bad fit for noisy data can be hard to detect. Reducing the dimensionality of the data via score compression, we investigate whether it would be easier for any bad fit to be spotted. Hence, given the results presented in Sect. 6.3, we test the robustness of the method against unmodelled effects in the correlation functions, via the χ^2 statistics.

[4] Note also that this kind of analysis heavily relies on mocks being consistent with each other (both in terms of mock production, and in terms of analysis), in order to avoid introducing extra variance.

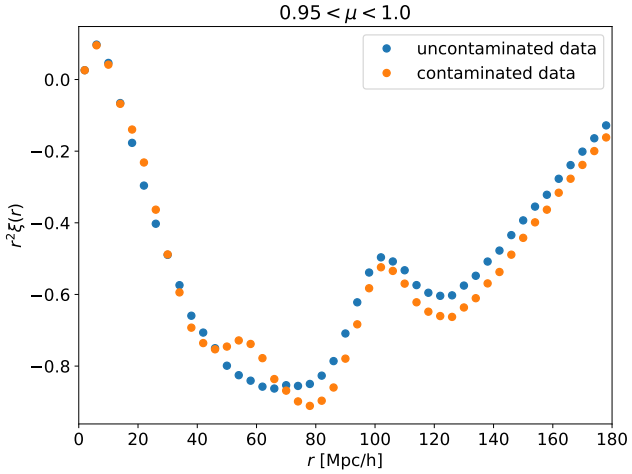

Figure 6.4.2: This wedge plot, for $|\mu| = |r_\parallel/r|$ between 0.95 and 1.0, shows the effect of adding metals (in orange) to the correlation model $\boldsymbol{\xi}$ without metals (in blue) along the line-of-sight. For simplicity in the χ^2 analysis we do not include contamination coming from HCD, so these features are only the effects of metal lines. Also, in this example, in order to better visualize the difference between the two, we have been generating noise from the covariance matrix of the *stacked* auto-correlation mock.

To this end we test the goodness of fit on contaminated data when metals are not modelled. For simplicity, here we restrict to the Lyα auto-correlation alone and without considering contamination from HCD. The sampled parameters will only be $\{\alpha_\parallel, \alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, \sigma_\parallel, \sigma_\perp\}$. Tests are run by constructing the χ^2 distributions over a set of 300 Monte Carlo realizations of the auto-correlation, introduced in Sect. 6.2.3: for each realization we run a minimizer and evaluate the χ^2 at the best fit.

We considered two main Monte Carlo populations: with and without metal contamination. The difference between the two is shown in the wedge plot of Fig. 6.4.2, which is built by averaging over the values of the correlation function in the 'wedge' of the space $\{r_\parallel, r_\perp\}$ identified by values of $|\mu| = |r_\parallel/r|$ between 0.95 and 1.0. To generate them we used the best fit values of $\{\alpha_\parallel, \alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, \sigma_\parallel, \sigma_\perp, b_{\eta,\mathrm{SiII}(1260)}, b_{\eta,\mathrm{SiII}(1193)}, b_{\eta,\mathrm{SiIII}(1207)}, b_{\eta,\mathrm{SiII}(1190)}\}$ for the contaminated *stacked* Lyα mock auto-correlation, where depending on the population (contaminated or uncontaminated) the metals' parameters were either included or not.

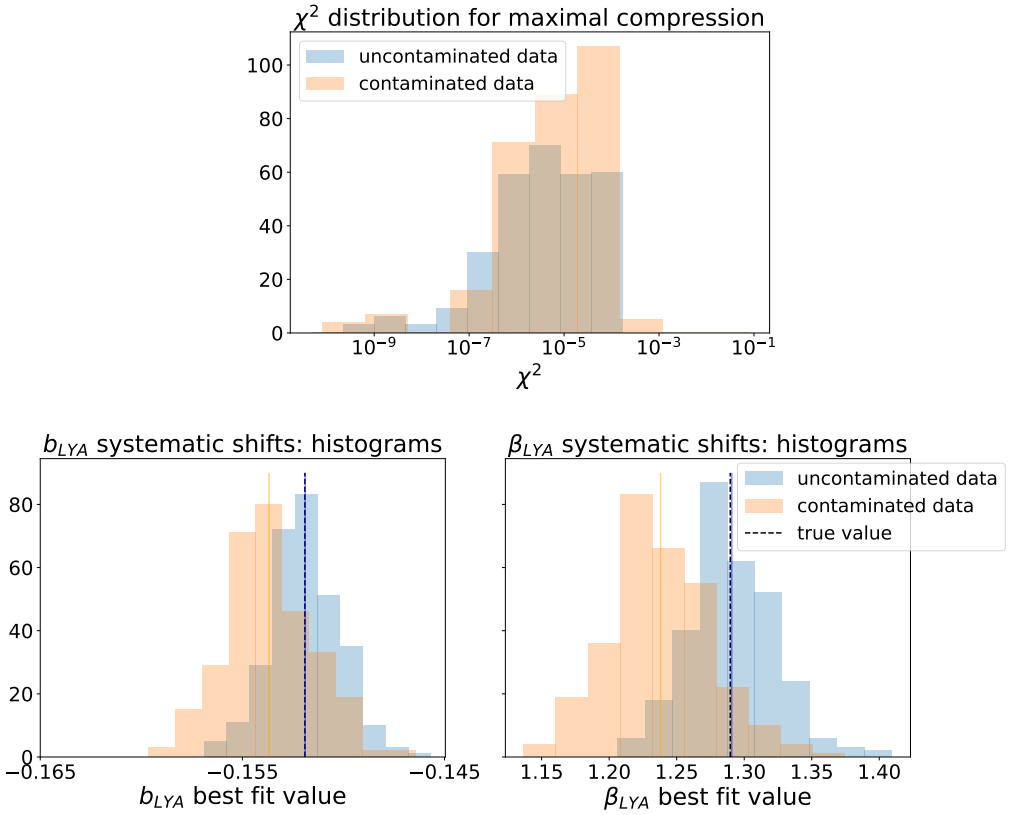

Figure 6.5.1: χ^2 histograms (left panel) for the *maximal* compression and corresponding best fit values histograms for the Lyα parameters (right panels), where blue refers to the uncontaminated case and orange to contaminated. In the *maximal* compression setup $\boldsymbol{t} = \boldsymbol{t}_{\rm max} = \{t_{\alpha_\parallel}, t_{\alpha_\perp}, t_{b_{\rm Ly\alpha}}, t_{\beta_{\rm Ly\alpha}}, t_{\sigma_\parallel}, t_{\sigma_\perp}\}$. The black dashed lines in the two panels on the right correspond to the true values used to generate the Monte Carlo realisations.

6.5.1 Maximal compression

For both the contaminated and uncontaminated mock data, we apply a compression down to the same number of sampled parameters without including contamination in the modelling, with the summary statistics thus given by $\boldsymbol{t}_{\rm max} = \{t_{\alpha_\parallel}, t_{\alpha_\perp}, t_{b_{\rm Ly\alpha}}, t_{\beta_{\rm Ly\alpha}}, t_{\sigma_\parallel}, t_{\sigma_\perp}\}$. This is defined as *maximal compression*. In what follows we are interested in learning about the χ^2 distribution for the two Monte Carlo populations.

We found that for both contaminated and uncontaminated data, the χ^2 distributions are similar, with values of the order of $\mathcal{O}(10^{-10} - 10^{-3})$ (left panel of Fig. 6.5.1). However, comparing the fits to the contaminated and uncontaminated data, the best-fit parameter values are systematically shifted for some parameters. The distributions of the best-fit values for $b_{\rm Ly\alpha}$ and $\beta_{\rm Ly\alpha}$ are shown in the right panels of Fig. 6.5.1: for the fits to contaminated data, 80% and 90% of the best-fit values respectively for each parameter are below the true value.

The χ^2 values remain very small for the fits to contaminated data, which indicates that in the compressed space, the model without contaminants still has enough flexibility to perfectly fit the data: the system has zero degrees of freedom, given that we are sampling six parameters, and the compressed data vector has six components. Instead of the mismatch between the model without contaminants and the contaminated data being visible in the form of large χ^2 values, it is manifest through a systematic shift in the recovered parameter values from the truth, which in a realistic data fitting scenario could not be detected. This is linked to the fact that we are very close to a linear model scenario, meaning that in the compressed space the model still has enough flexibility to fit the data. This motivated a deeper testing of the framework, extending it to extra degrees of freedom as follows.

6.5.2 Non-maximal compression

Given the problem highlighted in the *maximal* framework, we tested the pipeline in a *non-maximal compression* case, where the extra degrees of freedom are given by the metals contaminating the data. Namely, the *maximal* summary statistics is now extended to include $\boldsymbol{t}_{\rm extra} = \{t_{b_{\eta,\rm SiII(1260)}},$

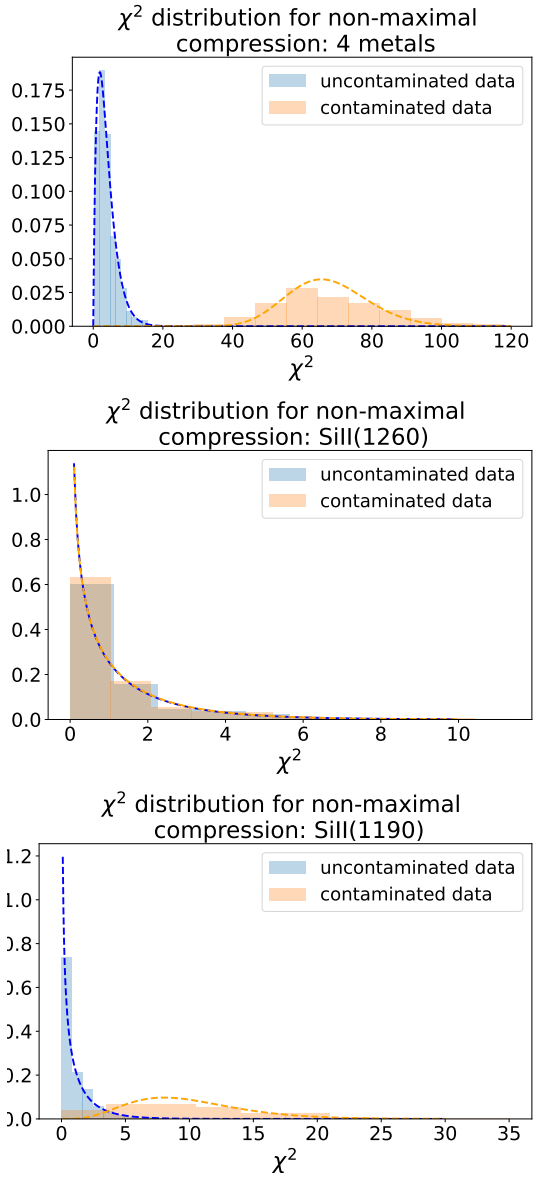

Figure 6.5.2: Normalized χ^2 histograms for the three *non-maximal compression* cases presented in Sect. 6.5.2: starting from the left, all four metals, SiII(1260) and SiII(1190) were used to build extra degrees of freedom. In blue the histograms and χ^2 distributions for the uncontaminated data, orange for contaminated. The corresponding χ^2 distributions (dashed lines) are generated assuming as number of degrees of freedom the mean of the histogram distributions. The first set of histograms, that relates to all four extra degrees of freedom, present a strong shift between the orange and the blue distributions: their corresponding means are 3.89 and 67.51. In the SiII(1260) case, both distributions have a mean of $\sim$ 1.1, while in the SiII(1190), the mean for the contaminated case is 1.01, against 10.04 in the contaminated case.

$t_{b_{\eta,\mathrm{SiII(1193)}}}, t_{b_{\eta,\mathrm{SiIII(1207)}}}, t_{b_{\eta,\mathrm{SiII(1190)}}}\}$. Still, metals will not be included in the likelihood modelling. This means that if the quantities of reference here are the compressed data vector

$$\boldsymbol{t} = \nabla\boldsymbol{\mu}_*^T \boldsymbol{C}^{-1}(\boldsymbol{d} - \boldsymbol{\mu}_*) \,, \tag{6.15}$$

the compressed model

$$\boldsymbol{\mu_t} = \nabla\boldsymbol{\mu}_*^T \boldsymbol{C}^{-1}(\boldsymbol{\mu(\theta)} - \boldsymbol{\mu}_*) \,, \tag{6.16}$$

and they enter the χ^2 as per

$$\chi^2(\boldsymbol{\theta}) = [\boldsymbol{t} - \boldsymbol{\mu_t}(\boldsymbol{\theta})]^T \boldsymbol{F}^{-1}[\boldsymbol{t} - \boldsymbol{\mu_t}(\boldsymbol{\theta})] \,, \tag{6.17}$$

the fiducial model $\boldsymbol{\mu}_*$ and its gradient will now include contaminants, whereas $\boldsymbol{\mu(\theta)}$ will not and $\boldsymbol{d}$ will either be contaminanted or uncontaminated data depending on the population used to build the χ^2 statistics. Now $\boldsymbol{t} = \{\boldsymbol{t}_{\mathrm{max}}, \boldsymbol{t}_{\mathrm{extra}}\}$. The length of the compressed data vector is ten, where the first six components refer to the sampled parameters, with a remainder of four components, which are fixed and constitute our extra degrees of freedom. Under the approximation that the mean of a χ^2 distribution indicates the number of degrees of freedom of the problem, we would expect that mean to be at least equal to the number of extra degrees of freedom we added. In our case, we expect that for the uncontaminated case, for which we know the modelling is good, the mean will be close to 4 (four metals). We want to test whether in this case a bad fit to the contaminated data is apparent as a mean χ^2 significantly larger than 4.

The χ^2 histograms are shown in the left panel of Fig. 6.5.2: the mean values for the uncontaminated and contaminated cases are respectively 3.89 and 67.51. Considering a χ^2 with number of degrees of freedom equal to 4, the p-values for the two means are respectively 0.4 and 10^{-14}: the bad fit in the contaminated case has emerged.

We further experimented over the addition of metals and we considered adding a single extra degree of freedom at a time, associated to either one of the following metals: the SiII(1260) and the

SiIII(1190). The resulting χ^2 histograms are shown in the middle and right panels of Fig. 6.5.2, respectively. These two metal lines were chosen because of how differently they affect the data: while the SiII(1260) contamination happens around the BAO scale along the line-of-sight, the SiII(1190) contributes to the peak at $\sim 60\mathrm{Mpc}/h$. We run the same exact experiment and find that the addition of $t_{b_{\eta,\mathrm{SiII}(1190)}}$ does bring out the bad fit, while the other does not. Specifically, the two χ^2 distributions when the extra degree of freedom is given by $b_{\eta,\mathrm{SiII}(1260)}$ have a mean of ~ 1, again equal to the number of degrees of freedom, but they cannot be distinguished. The p-values for both distributions, assuming one degree of freedom, are all above a threshold of 0.01. Both distributions are indicative of an acceptable fit. On the contrary, adding the extra compressed component related to SiII(1190) results in having a mean χ^2 of 1.01 in the uncontaminated case and 10.04 in the contaminated one, with corresponding p-values of 0.3 and 10^{-3} if we consider a target χ^2 distribution of one degree of freedom. This perhaps is indicative about the fact that in order to capture a bad fit, adding extra degrees of freedom is not enough: these extra degrees of freedom must be informative about features not captured by the core set of parameters. The SiII(1260) affects the model at scales of the correlation function which are on top of the BAO peak, which we model for, whereas SiII(1190) effectively adds information on a feature which is completely unmodelled.

In light of this, a possible solution is to add some extra degrees of freedom to the *maximal* compression vector, which are designed to be orthogonal to the already known components in the compressed space. This would allow the extra flexibility, that is not captured in the model, to highlight for a bad fit in the compressed framework. This is an interesting problem which is left for future work. However, a similar solution has already been implemented in the context of MOPED (Heavens et al., 2020), specifically to allow new physics to be discovered.

Not modelling the SiII(1260) line in the uncompressed traditional framework does not result in any bad fit, which makes this an example of systematics hidden in the large original data vector. At the same time, the fact that the SiII(1260) test in the compressed framework fails to show a bad fit at the level of the χ^2 is quite problematic, given this metal line is one of the primary contaminants

we have to be careful of in BAO measurement, affecting the peak's scale. The worry is then that, despite constructing an extended framework, there is a chance that some systematics hiding in the signal could be missed. This effectively means that in order to apply data compression, the underlying physics must be already well known to a good extent. Because some systematics could be either hard to model or to detect, in this example, we deliberately assumed we had no knowledge about known systematics, where in principle we could have also marginalized over them (Alsing and Wandelt, 2019b).

6.6 Robustness to parameter non-linearities

Each component of the score-compressed data vector relates to a specific model parameter, as per Eq. (6.8), via the gradient. Throughout the analysis, the BAO parameters proved to be a source of non-linearities in relation to their summary statistics components (see Fig. 6.2.1), sometimes resulting in a multi-peaked posterior distribution. With the intent of mitigating this effect, we were forced to impose a tight prior on both $\{\alpha_{\parallel}, \alpha_{\perp}\}$, which reduces the generalizability of the approach.

Based on the work of Protopapas et al. (2005), we explore extensions to the algorithm by considering an ensemble of fiducial values of the BAO parameters to compute the score-compressed vector components related to $\{\alpha_{\parallel}, \alpha_{\perp}\}$. For any extra set of BAO parameters $\{\alpha_{\parallel}^{\rm extra}, \alpha_{\perp}^{\rm extra}\}$, we introduce two extra summary statistics components:

$$\boldsymbol{t}_{\alpha_{\parallel}}^{\rm extra} = \nabla_{\alpha_{\parallel}} \boldsymbol{\mu}_{\rm extra}^{T} \boldsymbol{C}^{-1} (\boldsymbol{d} - \boldsymbol{\mu}_{\rm extra}) \,, \tag{6.18}$$

$$\boldsymbol{t}_{\alpha_{\perp}}^{\rm extra} = \nabla_{\alpha_{\perp}} \boldsymbol{\mu}_{\rm extra}^{T} \boldsymbol{C}^{-1} (\boldsymbol{d} - \boldsymbol{\mu}_{\rm extra}) \,, \tag{6.19}$$

where $\boldsymbol{\mu}_{\rm extra}$ is the model evaluated at $\{\boldsymbol{\alpha}_{\parallel}^{\rm extra}, \boldsymbol{\alpha}_{\perp}^{\rm extra}\}$, keeping the previously defined fiducial values for the other parameters. As these extra components effectively represent an extension of the compressed dataset, the Fisher matrix in Eq. (6.10) will also be expanded to include $[\nabla_{\alpha_{\parallel,\perp}} \boldsymbol{\mu}_{\rm extra}]^{T} \boldsymbol{C}^{-1} [\nabla_{\alpha_{\parallel,\perp}}^{T} \boldsymbol{\mu}_{\rm extra}]$. We test this extension on the same mock that was used to

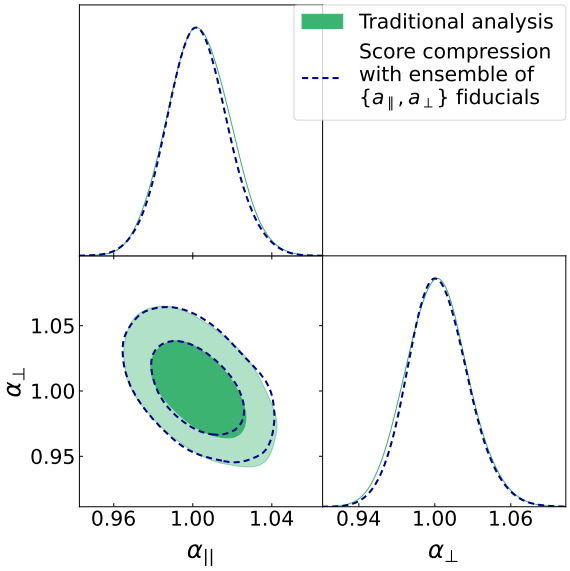

Figure 6.6.1: Triangle plots of the BAO parameters of interest $\{\alpha_{\parallel}, \alpha_{\perp}\}$ for one set of the Lyα auto- and cross- mock correlations, with relaxed priors. The green contours refer to the results obtained performing the inference using the full uncompressed data vector, which we denote as 'Traditional analysis', while the blue dashed refer to the compressed analysis results, denoted as 'Score compression analysis'. The framework of the latter is extended here to the assumption of multiple fiducial values for $\{\alpha_{\parallel}, \alpha_{\perp}\}$ when performing the compression, namely $[\{\alpha_{\parallel} = 1.00, \alpha_{\perp} = 1.01\}, \{\alpha_{\parallel} = 0.8, \alpha_{\perp} = 1.2\}, \{\alpha_{\parallel} = 1.2, \alpha_{\perp} = 0.8\}, \{\alpha_{\parallel} = 1.3, \alpha_{\perp} = 0.7\}, \{\alpha_{\parallel} = 0.9, \alpha_{\perp} = 1.1\}]$.

test the subsampling covariance matrix in Sect. 6.4, and results are presented in Fig. 6.6.1, imposing a physically motivated uniform prior $[0.65, 1.35]$ for both $\alpha_{\parallel}$ and $\alpha_{\perp}$. The ensemble of extra fiducials is given by the set $[\{\alpha_{\parallel} = 0.8, \alpha_{\perp} = 1.2\}, \{\alpha_{\parallel} = 1.2, \alpha_{\perp} = 0.8\}, \{\alpha_{\parallel} = 1.3, \alpha_{\perp} = 0.7\}, \{\alpha_{\parallel} = 0.9, \alpha_{\perp} = 1.1\}]$, in addition to the original $\{\alpha_{\parallel} = 1.00, \alpha_{\perp} = 1.01\}$ (see Tab. 6.2.1). From Fig. 6.6.1 it can be seen that the constraining power on the BAO parameters between the traditional and compressed methods match. This same result is also true for the other parameters, not shown here.

We tested the extension in terms of generalizability by progressively adding extra points to the ensemble, with reasonable spread, and found that with an ensemble of three to four extra fiducial sets of BAO parameters the algorithm is able to effectively get rid of the secondary posterior peaks and increase the accuracy of the measurement. Hence, the assumption of multiple fiducials

for the BAO parameters, for which we had to impose a tight prior, enables us to relax the prior constraints.

6.7 Application to real data

The score compression framework has so far been tested on realistic mocks, hence it is straightforward to apply this same algorithm to real eBOSS DR16 Lyα data, for which we refer to du Mas des Bourboux et al. (2020). The set of nuisance parameters is now extended to also include the contamination from carbon absorbers, the systematic quasar redshift error $\Delta r_{\parallel}$, the quasar radiation strength ξ_0^{TP} and the sky-subtraction parameters $A_{\mathrm{sky,Ly}\alpha}$ and $\sigma_{\mathrm{sky,Ly}\alpha}$. The results presented in Sect. 6.6 motivate a direct test of the whole extended framework, which gets rid of the tight prior, to the real data. The ensemble of BAO parameter fiducial values is given by the set of $\{\alpha_{\parallel} = 1.05, \alpha_{\perp} = 0.96\}$ — which are the best fit values obtained through the traditional analysis — and $[\{\alpha_{\parallel} = 0.8, \alpha_{\perp} = 1.2\}, \{\alpha_{\parallel} = 1.2, \alpha_{\perp} = 0.8\}, \{\alpha_{\parallel} = 1.3, \alpha_{\perp} = 0.7\}, \{\alpha_{\parallel} = 0.9, \alpha_{\perp} = 1.1\}]$, which were found to be effective in Sect. 6.6. The fiducial values of the other parameters are set to the best fit found with the standard uncompressed analysis. In Fig. 6.7.1, we present the agreement of the extended framework against the traditional approach at the level of $\{\alpha_{\parallel}, \alpha_{\perp}, b_{\eta,\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, \Delta r_{\parallel}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}\}$. The nuisance parameters are also found to be in excellent agreement.

6.8 Conclusions

Standard analyses of the Lyman-α (Lyα) forest correlation functions focus on a well localized region, which corresponds to the baryon acoustic oscillations (BAO) peak. However, these correlation functions usually have dimensions of 2500 or 5000, which means the cosmological signal is extracted from a small subset of bins. This means that reducing the dimensionality of the data vector, while retaining the information we care about, could be a step forward in optimizing the analysis. At the same time, as extensively explained in Sect. 6.2, the covariance matrix $\mathbf{C}$ used

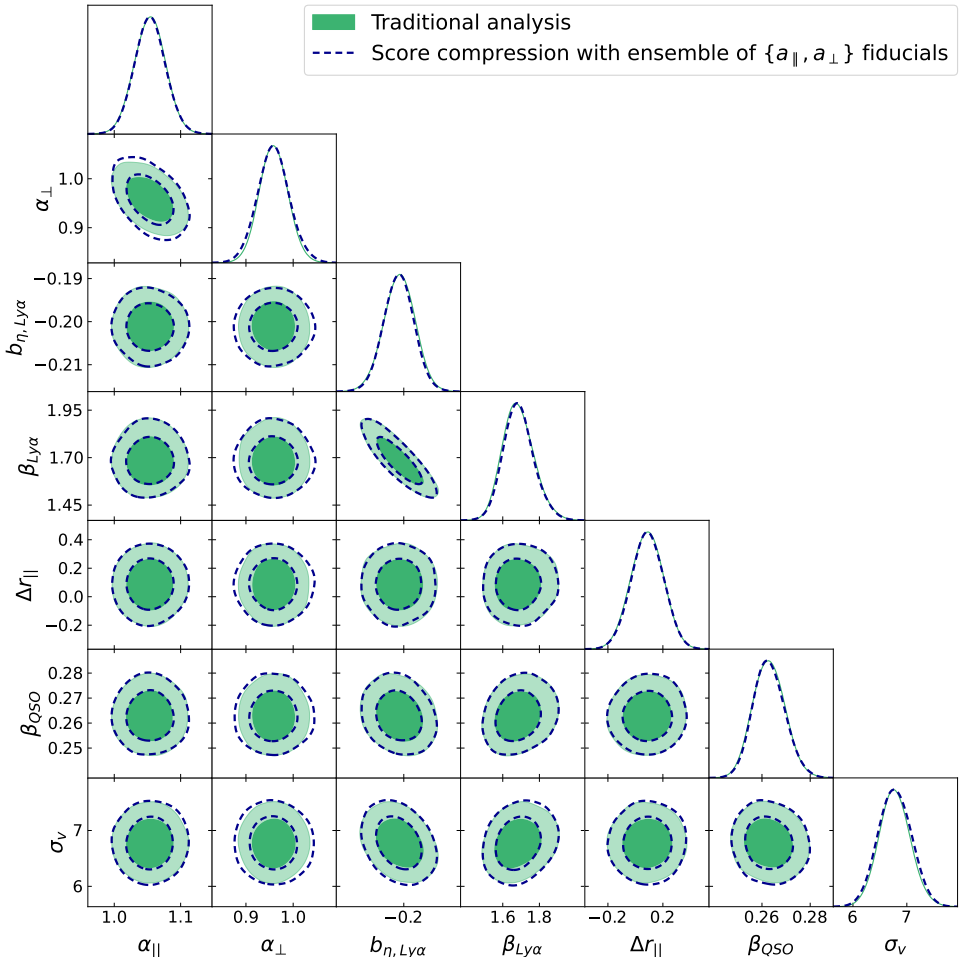

Figure 6.7.1: Triangle plot for fits to the real eBOSS DR16 data Lyα auto- and cross- correlations, using the traditional approach (green) and the score compression framework (dashed blue) extended to include extra fiducial values of the BAO parameters at $[\{\alpha_\parallel = 0.8, \alpha_\perp = 1.2\}, \{\alpha_\parallel = 1.2, \alpha_\perp = 0.8\}, \{\alpha_\parallel = 1.3, \alpha_\perp = 0.7\}, \{\alpha_\parallel = 0.9, \alpha_\perp = 1.1\}]$. The results shown here are for the standard parameters $\{\alpha_\parallel, \alpha_\perp, b_{\eta,\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, \Delta r_\parallel, \beta_{\mathrm{QSO}}, \sigma_\mathrm{v}\}$.

for Lyα correlations analyses is estimated via sub-sampling. However, the dimensionality of the correlation functions is much larger than the number of data samples used to estimate the covariance. Reducing the dimensionality of the data vector to $\mathcal{O}(10)$ allows for a reliable estimate of the covariance matrix. Given these premises, the goal of this work is to apply and explore a data compression algorithm for realistic Lyα auto- and cross-correlation functions.

We reduced the dimensionality of the data vector to a set of summary statistics $\mathbf{t}$ using score compression. We assume a Gaussian likelihood, test for its validity, and show that this assumption is preserved in the compressed space as well, as the compression is a linear transformation. In the compressed space the covariance can be either given by the mapped traditional covariance or by a covariance estimated directly in such a space.

We tested the compressed framework against the traditional approach at the posterior level, when using the original covariance $\mathbf{C}$, and found the two of them agree, and no bias is introduced. We then showcased a test example of covariance matrix evaluation in the compressed space, which is a key benefit of the approach, enabling a comparison to the covariance matrix obtained in the traditional sub-optimal framework. Because of non-linear relationship between the BAO parameters and their summary statistics components, throughout the analysis we adopted a tight prior on $\{\alpha_{\parallel}, \alpha_{\perp}\}$. Later in the analysis, with the aim of increasing the generalizability of the approach, while relaxing the prior constraint, we successfully tested extensions to the framework by assuming an ensemble of fiducial values for these problematic parameters.

We then further examined the compressed framework, by testing the inference against unmodelled effects and we find that if any information about the unmodelled features in the correlation function is not captured by the compressed data vector $\boldsymbol{t}$, this can potentially lead to biases, which do not emerge at the level of the χ^2 goodness of fit test. Hence, we advise against performing goodness of fit tests in compressed space, unless the compressed vector is extended to include extra degrees of freedom, analogous to what is done in Heavens et al. (2020). Extending the framework in this sense is left for future work.

We applied our extended compression framework to DR16 data from the Extended Baryon Oscillation Spectroscopic Survey and demonstrated that the posterior constraints are accurately recovered without loss of information. A step change in constraining power, and thus accuracy requirements, is expected for forthcoming Lyα cosmology analyses by the on-going DESI experiment (see e.g., Gordon et al., 2023), which will observe up to 1 million high-redshift quasars with $z > 2$. Optimal data compression as proposed in this work will facilitate these analyses through inference that is complementary to the traditional approach and through additional consistency and validation tests.

6.9 Appendix: Full results for the mock to mock covariance test

We here present in Fig. 6.9.1 the full set of results from the mock-to-mock covariance test, presented in Sect. 6.4, against the contours obtained using the original covariance in the compressed framework. Numerical values are reported in Tab. 6.2.1. The contours agree well with each other. The most striking enlargements of the posteriors are visible for the parameters $\{\alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{HCD}}\}$. Because the 'Original covariance' setup has been shown to agree with the standard analysis in Sect. 6.3, this comparison automatically becomes a comparison to the standard approach.

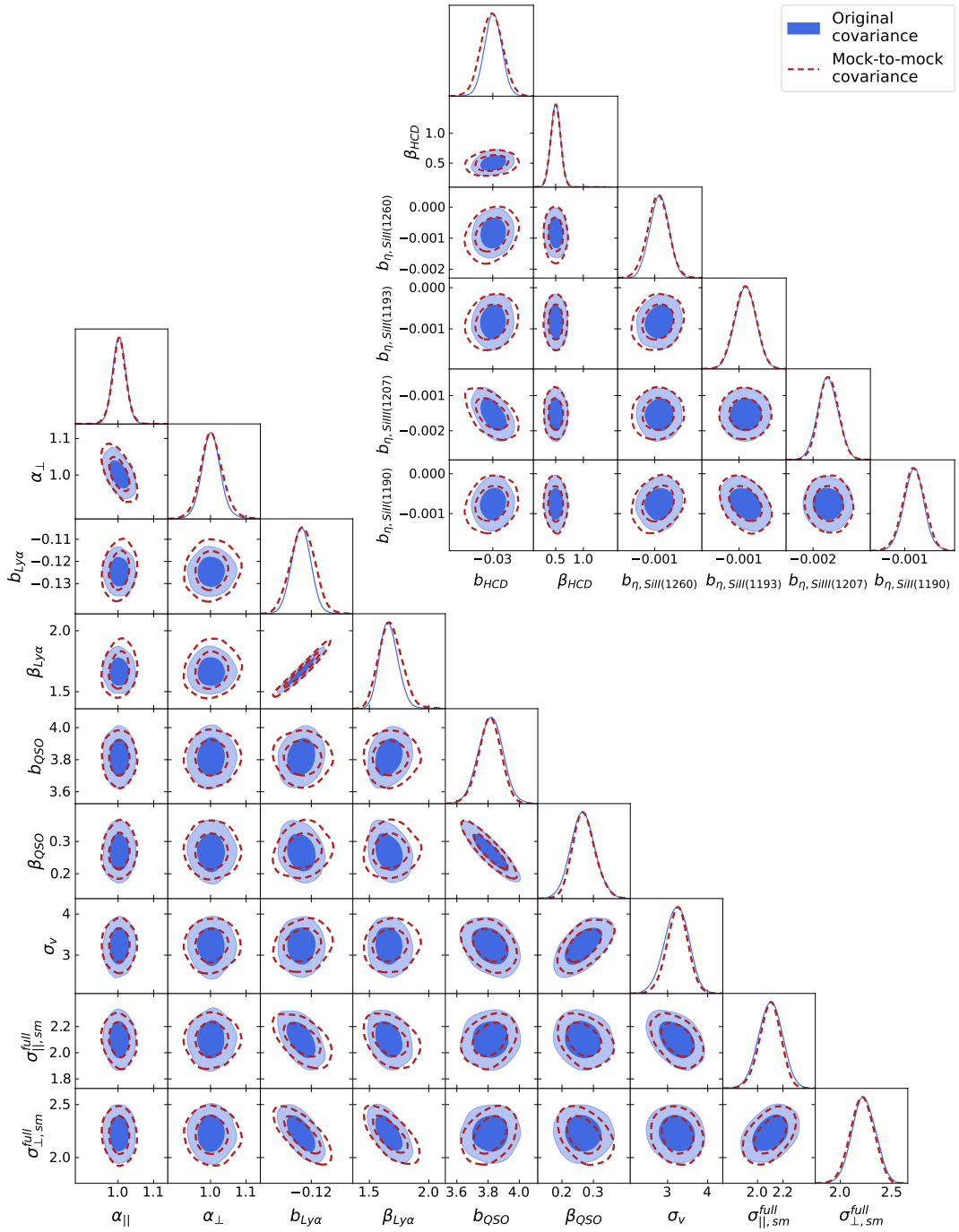

Figure 6.9.1: Triangle plots of the parameters of interest for one set of the Lyα auto- and cross- correlation mocks. Results are split, for presentation purposes only, into the set of standard parameters $\{\alpha_\parallel, \alpha_\perp, b_{\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\mathrm{QSO}}, \beta_{\mathrm{QSO}}, \sigma_{\mathrm{v}}, \sigma_\parallel, \sigma_\perp\}$ (lower left panel) and contaminants parameters $\{b_{\eta,\mathrm{SiII}(1260)}, b_{\eta,\mathrm{SiII}(1193)}, b_{\eta,\mathrm{SiIII}(1207)}, b_{\eta,\mathrm{SiII}(1190)}, b_{\mathrm{HCD}}, \beta_{\mathrm{HCD}}\}$ (upper right panel). The blue contours refer to the results obtained performing the inference using the original covariance matrix $\boldsymbol{C}$ mapped into the compressed space (the Fisher matrix) in the likelihood function, and hence are denoted as 'Original covariance'. On the other hand, the red dashed results, denoted as 'Mock-to-mock covariance', refer to the case in which the mock-to-mock covariance matrix is used instead, while adopting a t-distribution likelihood.

Part III

Conclusions

7 Conclusions and outlook

The base assumptions of the standard model have been proven to be robust over a broad range of observables, however there is still no clarity over the nature of some constituents of the Universe, such as dark energy. At the same time, our current belief of what this component should be is challenged by arising tensions, the most famous one being related to the rate of expansion at the present time, H_0 (Di Valentino et al., 2021). Increasing the variety, accuracy and size of available probes should shed light on what is driving this tension, whether that is systematics or deviations from our current model.

In light of this, inferring cosmological parameters is becoming incredibly challenging: not only the size of the datasets is increasing, but the modelling is becoming more and more complex. This means that the joint space of data and parameters is becoming unfeasible to explore with traditional methods. It is then crucial to explore more efficient approaches, that involve data compression and forward-modelling.

This thesis work explored some applications of these methods, in the field of gravitational waves (GW) and Lyman-α (Lyα) forest.

7.1 Simulation-based inference in GW analyses

The last decade marked the beginning of a series of successful observing runs for the LIGO-Virgo-KAGRA (LVK) Collaboration, which just started the fourth one in May 2023. The upgrades to the facilities are leading to the observation of further and fainter systems (Abbott et al., 2020). Such advancements challenges data pipelines, both from the detection and parameter inference perspectives. Rapidly identifying the sources of gravitational waves to allow for efficient multi-messenger astronomy is as important as building faster, yet reliable, inference frameworks. With this in mind, the last couple of years have seen the effort, among the GW community, in develop-

ing inference tools based on forward-modelling (Green and Gair, 2021; Dax et al., 2021; Bhardwaj et al., 2023).

7.1.1 Summary: unbiased inference

Motivated by numerical challenges of traditional analyses and by the interest in the Hubble constant, in Chapter 4 we tested the reliability of such a simulation-based framework against the Malmquist bias, due to selection effects, while performing population level inference of H_0 from standard sirens. The main advantage of using forward-modelling in the inference pipeline is that selection is naturally built into the simulations. By constructing a hierarchical toy model, we were able to run both simulation-based inference (SBI) and Markov chain Monte Carlo (MCMC) frameworks, finally comparing the precision and accuracy achievable using this alternative method in the presence of GW selection effects alone. We used the pydelfi and pystan tools respectively for the two approaches. As we have seen in Chapter 3, the inputs to the neural density estimators (NDE), used within SBI, must be a set of low-dimensional summary statistics, which then requires the choice of a compression scheme to be applied to the binary neutron star mergers catalogue. To solve this task we used a simple regression neural network. As a general note, this means that not only does the performance of the SBI framework largely depend on the modelling, but also on the ability of the compression method to preserve fundamental information we wish to infer. In this proof-of-concept work, we demonstrated that the SBI framework does not introduce any bias in the presence of GW selection effects, being as precise as the full Bayesian hierarchical model sampling. Marginalising over the knowledge of the ground truth increases the H_0 uncertainty by only 6% for training sets of the regression neural network consisting of $O(10^4)$ populations. Interestingly, by augmenting the size of the training set, the uncertainty increase drops.

7.1.2 Extended population inference pipelines

The first natural extension to the framework we have presented here is the addition of EM selection effects. As the goal of the work was indeed to test Hubble constant estimation against the

Malmquist bias, only the GW selection effect was taken into account. However, to reach a direct estimate of the redshift of the source, an EM counterpart must be observed, and that signal is subject to selection effects too. The work by Gagnon-Hartman et al. (2023) represents an extension to our analysis in this sense.

We have been testing simulation-based inference at a population-level, while performing cosmological parameter estimation, with a particular focus on the Hubble constant. However, the interest over gravitational waves signal is not limited to what we can learn about cosmology. Indeed, these phenomena are also informative about the physics of binary neutron star systems themselves, in regards to the mass distributions and their equation of state. Soon after the publication of this work, there was an effort on this side in works such as those of Veske et al. (2021) and Golomb and Talbot (2022).

Extensions to the simulation-based frameworks, to model the complexity of the studied systems and systematics, are crucial to make sure that our pipelines will be reliable by the time third-generation GW detectors such as Einstein Telescope (Sathyaprakash et al., 2012) and Cosmic Explorer (Abbott et al., 2017) come online.

7.2 Encoding and extracting information from Lyα correlation functions

The scale of the baryon acoustic oscillation (BAO) is a well established probe for cosmology: it probes the matter density field over a wide range of redshifts, successfully constraining the expansion history (Seo and Eisenstein, 2003). This same feature is also detected in Lyα forest correlation functions, being the absorption lines tracer of matter along the line of sight, and it probes redshifts between 2 and ~ 4.

When performing cosmological inference, standard analyses of Lyα correlation functions only consider the information carried by this distinct peak, neglecting the information coming from

the broader range of scales. In May 2021, the Dark Energy Spectroscopic Instrument (DESI) completed the Survey Validation and recently published EDR results (DESI Collaboration et al., 2023). As data from the DESI Main Survey will become available, it is essential to test and understand whether:

- this peak encodes all the relevant cosmological information ;
- this same information can be compressed down to optimise the inference pipeline.

7.2.1 Summary: inferring cosmology

Following up the work of Cuceu et al. (2021), which highlights the importance of considering the broadband components, we are motivated in performing a direct fit to the full shape, in order to address the first point. This work has been presented in Chapter 5. We extended the vega framework to allow for cosmological parameters sampling through cobaya, for which we implemented an *ad-hoc* Gaussian likelihood. We tested the framework on synthetic idealised Lyα correlation functions and provided a comparison to the classic method, while sampling $\{H_0, \Omega_\mathrm{M}, \Omega_\mathrm{B}h^2, A_\mathrm{s}, n_\mathrm{s}\}$ and marginalising over Lyα and quasar model parameters. We obtain better constraints on cosmological parameters, and, by extending the comparison to the analysis of Cuceu et al. (2021), we find that most of the extra information can be traced back to the Alcock-Paczyński effect and redshift space distortions.

7.2.2 Summary: data compression

Two-dimensional Lyα correlation functions have a few thousand data points, whereas the BAO peak is only limited to a small region. If BAO is the only feature from which we wish to extract cosmological information, the analysis could be largely optimised by compressing the data down, preserving the information for which we care. In Chapter 6 we applied score compression to realistic Lyα auto- and cross- correlation functions, under the assumption of a Gaussian likelihood,

which we test for. Compressing the data vector comes with an additional advantage. The constraining power on sampled parameters is tightly linked to the covariance matrix of the correlation functions, which however can only be estimated from data. While in standard analyses this is not true, in the compressed framework the length of the data vector is lower than the number of samples used to estimate the covariance matrix, which means the latter can be reliably computed. This means that the secondary advantage of this framework is that it gives us the opportunity to test the accuracy of the covariance as computed in standard analyses. At the posterior level, we compare the results obtained using the traditional approach against the compressed one, finding that no bias is introduced and the original estimate of the covariance matrix through the subsampling method is good enough.

7.2.3 Extensions to the frameworks

Both the works presented in Chapter 5 and Chapter 6 extended current Lyα forest frameworks, to either understand what information can be extracted from the correlation functions or if that can be compressed down to a set of summary statistics. The fact that these two works are complementary means that a natural extension of both would be computing the score-compressed vector by directly encoding the cosmological information. This would effectively mean computing derivatives of the log-likelihood with respect to the parameters $\{H_0, \Omega_{\rm M}, \Omega_{\rm B}h^2, A_{\rm s}, n_{\rm s}\}$ sampled in the first analysis and build a new summary statistics still based on score compression.

However, to reach such a goal, the framework built in Chapter 5 should be tested against systematics, while including more realistic features in the data, such as the effects of contaminants, which were instead included in the compressed analysis.

In light of this, future work could then perform cosmological inference with the aid of score compression on DESI data, which are now starting to become available. At the same time, real data analyses might reflect physics not included in the current modelling and, as we have seen in Chapter 6, the current score compression algorithm does not deal well with model mismatch at the level of the χ^2. For this reason, before doing such a step, it is necessary to extend the mapping

to the compressed space as in Heavens et al. (2020). This would mean extending the summary statistics to orthogonal components which are informative about unknown effects.

This concludes the research I carried out during my PhD. This work is set in the context of testing novel techniques for data compression and simulation-based inference, with the scope of optimising cosmological analyses, yet making sure results are reliable. Over the last few years, with the help of Machine Learning, there has been a boost in the development and usage of these approaches, which will be key to perform inference in the upcoming years, given that the data sample size and the model complexity are increasing.

Bibliography

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". *Phys. Rev. Lett.*, 116:061102, February 2016. doi:10.1103/PhysRevLett.116.061102.

B P Abbott, R Abbott, T D Abbott, M R Abernathy, K Ackley, C Adams, P Addesso, R X Adhikari, V B Adya, C Affeldt, et al. "Exploring the sensitivity of next generation gravitational wave detectors". *Class. Quantum Grav.*, 34(4):044001, January 2017. doi:10.1088/1361-6382/aa51f4.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. "A gravitational-wave standard siren measurement of the Hubble constant". *Nature*, 551(7678):85–88, October 2017. doi:10.1038/nature24471.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. "Multi-messenger Observations of a Binary Neutron Star Merger". *ApJL*, 848(2):L12, October 2017a. doi:10.3847/2041-8213/aa91c9.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral". *Phys. Rev. Lett.*, 119:161101, October 2017b. doi:10.1103/PhysRevLett.119.161101.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. "GW170817: Measurements of Neutron Star Radii and Equation of State". *Phys. Rev. Lett.*, 121(16), October 2018. doi:10.1103/physrevlett.121.161101.

B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, et al. "Prospects for observing and localizing gravitational-wave

transients with Advanced LIGO, Advanced Virgo and KAGRA". *Living Reviews in Relativity*, 23(1), September 2020. doi:10.1007/s41114-020-00026-9.

T. M. C. Abbott, M. Aguena, A. Alarcon, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, E. Baxter, et al. "Dark Energy Survey Year 3 results: Constraints on extensions to ΛCDM with weak lensing and galaxy clustering". *Phys. Rev. D*, 107(8), April 2023. doi:10.1103/physrevd.107.083504.

G. E. Addison, D. J. Watts, C. L. Bennett, M. Halpern, G. Hinshaw, and J. L. Weiland. "Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy". *ApJ*, 853(2):119, January 2018. doi:10.3847/1538-4357/aaa1ed.

Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A. Blazek, Adam S. Bolton, Joel R. Brownstein, Angela Burden, Chia-Hsun Chuang, et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample". *MNRAS*, 470(3):2617–2652, March 2017. doi:10.1093/mnras/stx721.

Shadab Alam, Marie Aubert, Santiago Avila, Christophe Balland, Julian E. Bautista, Matthew A. Bershady, Dmitry Bizyaev, Michael R. Blanton, Adam S. Bolton, Jo Bovy, et al. "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". *Phys. Rev. D*, 103(8), April 2021. doi:10.1103/physrevd.103.083533.

C. Alcock and B. Paczynski. "An evolution free test for non-zero cosmological constant". *Nature*, 281:358–359, 1979. doi:10.1038/281358a0.

Justin Alsing and Benjamin Wandelt. "Generalized massive optimal data compression". *MNRAS: Letters*, 476(1):L60–L64, 02 2018. ISSN 1745-3925. doi:10.1093/mnrasl/sly029.

Justin Alsing and Benjamin Wandelt. "Nuisance hardened data compression for fast likelihood-free inference". *MNRAS*, 488(4):5093–5103, July 2019a. doi:10.1093/mnras/stz1900.

Justin Alsing and Benjamin Wandelt. "Nuisance hardened data compression for fast likelihood-free inference". *MNRAS*, 488(4):5093–5103, 07 2019b. ISSN 0035-8711. doi:10.1093/mnras/stz1900.

Justin Alsing, Benjamin Wandelt, and Stephen Feeney. "Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology". *MNRAS*, 477(3): 2874–2885, March 2018. doi:10.1093/mnras/sty819.

Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. "Fast likelihood-free cosmology with neural density estimators and active learning". *MNRAS*, July 2019. doi:10.1093/mnras/stz1960.

Andreu Arinyo-i-Prats, Jordi Miralda-Escudé, Matteo Viel, and Renyue Cen. "The non-linear power spectrum of the Lyman alpha forest". *JCAP*, 2015(12):017–017, December 2015. doi:10.1088/1475-7516/2015/12/017.

Éric Aubourg, Stephen Bailey, Julian E. Bautista, Florian Beutler, Vaishali Bhardwaj, Dmitry Bizyaev, Michael Blanton, Michael Blomqvist, Adam S. Bolton, Jo Bovy, et al. "Cosmological implications of baryon acoustic oscillation measurements". *Phys. Rev. D*, 92:123516, December 2015. doi:10.1103/PhysRevD.92.123516.

John N. Bahcall and E. E. Salpeter. "On the Interaction of Radiation from Distant Sources with the Intervening Medium". *ApJ*, 142:1677–1680, November 1965. doi:10.1086/148460.

M. Bailes, B. K. Berger, P. R. Brady, M. Branchesi, K. Danzmann, M. Evans, K. Holley-Bockelmann, B. R. Iyer, T. Kajita, S. Katsanevas, et al. "Gravitational-wave physics and astronomy in the 2020s and 2030s". *Nature Reviews Physics*, 3(5):344–366, January 2021. doi:10.1038/s42254-021-00303-8.

J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. "The Statistics of Peaks of Gaussian Random Fields". *ApJ*, 304:15, May 1986. doi:10.1086/164143.

Julian E. Bautista, Nicolás G. Busca, Julien Guy, James Rich, Michael Blomqvist, Hélion du Mas des Bourboux, Matthew M. Pieri, Andreu Font-Ribera, Stephen Bailey, Timothée Delubac, et al. "Measurement of baryon acoustic oscillation correlations at $z = 2.3$ with SDSS DR12 Lyα-Forests". *A&A*, 603:A12, June 2017. doi:10.1051/0004-6361/201730533.

Mark A Beaumont, Wenyang Zhang, and David J Balding. "Approximate Bayesian Computation in Population Genetics". *Genetics*, 162(4):2025–2035, 12 2002. ISSN 1943-2631. doi:10.1093/genetics/162.4.2025.

Richard Bellman and Robert Kalaba. "On adaptive control processes". *IRE Transactions on Automatic Control*, 4(2):1–9, 1959. URL https://ieeexplore.ieee.org/document/1104847.

Richard E Bellman. *Dynamic programming*. Princeton university press, 1957.

Jose Luis Bernal, Licia Verde, and Adam G. Riess. "The trouble with H_0". *JCAP*, 10:019, 2016. doi:10.1088/1475-7516/2016/10/019.

José Luis Bernal, Licia Verde, Raul Jimenez, Marc Kamionkowski, David Valcin, and Benjamin D. Wandelt. "The trouble beyond H_0 and the new cosmic triangles". *Phys. Rev. D*, 103(10), May 2021. doi:10.1103/physrevd.103.103533.

Michael Betancourt. "A Conceptual Introduction to Hamiltonian Monte Carlo", 2018. arXiv:1701.02434 [stat.ME].

Uddipta Bhardwaj, James Alvey, Benjamin Kurt Miller, Samaya Nissanke, and Christoph Weniger. "Peregrine: Sequential simulation-based inference for gravitational wave signals", 2023. arXiv:2304.02035 [gr-qc].

S. Birrer, T. Treu, C. E. Rusu, V. Bonvin, C. D. Fassnacht, J. H. H. Chan, A. Agnello, A. J. Shajib, G. C.-F. Chen, M. Auger, et al. "H0LiCOW – IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant". *MNRAS*, 484(4):4726–4753, January 2019. doi:10.1093/mnras/stz200.

Christopher M. Bishop. Mixture density networks. Technical report, 1994.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.

Michael Blomqvist, Hélion du Mas des Bourboux, Nicolás G. Busca, Victoria de Sainte Agathe, James Rich, Christophe Balland, Julian E. Bautista, Kyle Dawson, Andreu Font-Ribera, et al. "Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14". *A&A*, 629:A86, September 2019. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/201935641.

Samuel Brieden, Héctor Gil-Marín, and Licia Verde. "ShapeFit: extracting the power spectrum shape information in galaxy surveys beyond BAO and RSD". *JCAP*, 2021(12):054, December 2021. doi:10.1088/1475-7516/2021/12/054.

Dillon Brout and Daniel Scolnic. "It's Dust: Solving the Mysteries of the Intrinsic Scatter and Host-galaxy Dependence of Standardized Type Ia Supernova Brightnesses". *ApJ*, 909(1):26, March 2021. doi:10.3847/1538-4357/abd69b.

Dillon Brout, Dan Scolnic, Brodie Popovic, Adam G. Riess, Anthony Carr, Joe Zuntz, Rick Kessler, Tamara M. Davis, Samuel Hinton, David Jones, et al. "The Pantheon+ Analysis: Cosmological Constraints". *ApJ*, 938(2):110, October 2022. doi:10.3847/1538-4357/ac8e04.

N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby, J.-M. Le Goff, M. M. Pieri, A. Slosar, É. Aubourg, et al. "Baryon acoustic oscillations in the Lyα forest of BOSS quasars". *A&A*, 552:A96, April 2013. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/201220724.

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. "Stan: A probabilistic programming language". *Journal of Statistical Software*, 76(1), 2017. doi:10.18637/jss.v076.i01.

Tom Charnock, Guilhem Lavaux, and Benjamin D. Wandelt. "Automatic physical inference with information maximizing neural networks". *Phys. Rev. D*, 97(8), April 2018. doi:10.1103/physrevd.97.083004.

Hsin-Yu Chen. "Systematic Uncertainty of Standard Sirens from the Viewing Angle of Binary Neutron Star Inspirals". *Phys. Rev. Lett.*, 125(20), November 2020. doi:10.1103/physrevlett.125.201301.

Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. "A two per cent Hubble constant measurement from standard sirens within five years". *Nature*, 562(7728):545–547, 2018. doi:10.1038/s41586-018-0606-0.

Alvin J. K. Chua and Michele Vallisneri. "Learning Bayesian Posteriors with Neural Networks for Gravitational-Wave Inference". *Phys. Rev. Lett.*, 124:041102, January 2020. doi:10.1103/PhysRevLett.124.041102.

Shaun Cole, Will J. Percival, John A. Peacock, Peder Norberg, Carlton M. Baugh, Carlos S. Frenk, Ivan Baldry, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, et al. "The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications". *MNRAS*, 362:505–534, 2005. doi:10.1111/j.1365-2966.2005.09318.x.

P. S. Cowperthwaite, V. A. Villar, D. M. Scolnic, and E. Berger. "LSST Target-of-opportunity Observations of Gravitational-wave Events: Essential and Efficient". *ApJ*, 874(1):88, March 2019. doi:10.3847/1538-4357/ab07b6.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. "The frontier of simulation-based inference". *PNAS*, 117(48):30055–30062, 2020. doi:10.1073/pnas.1912789117.

Andrei Cuceu, James Farr, Pablo Lemos, and Andreu Font-Ribera. "Baryon Acoustic Oscillations and the Hubble constant: past, present and future". *JCAP*, 2019(10):044, October 2019. doi:10.1088/1475-7516/2019/10/044.

Andrei Cuceu, Andreu Font-Ribera, and Benjamin Joachimi. "Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman-α forest". *JCAP*, 2020(07):035–035, July 2020. doi:10.1088/1475-7516/2020/07/035.

Andrei Cuceu, Andreu Font-Ribera, Benjamin Joachimi, and Seshadri Nadathur. "Cosmology beyond BAO from the 3D distribution of the Lyman-α forest". *MNRAS*, 506(4):5439–5450, August 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab1999.

Andrei Cuceu, Andreu Font-Ribera, Paul Martini, Benjamin Joachimi, Seshadri Nadathur, James Rich, Alma X González-Morales, Hélion du Mas des Bourboux, and James Farr. "The Alcock–Paczyński effect from Lyman-α forest correlations: analysis validation with synthetic data". *MNRAS*, 523(3):3773–3790, June 2023. doi:10.1093/mnras/stad1546.

Andrei Cuceu, Andreu Font-Ribera, Seshadri Nadathur, Benjamin Joachimi, and Paul Martini. "Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman-α Forest". *American Physical Society (APS)*, 130(19):191003, May 2023. doi:10.1103/PhysRevLett.130.191003.

Neal Dalal, Daniel E. Holz, Scott A. Hughes, and Bhuvnesh Jain. "Short GRB and binary black hole standard sirens as a probe of dark energy". *Phys. Rev. D*, 74:063006, 2006. doi:10.1103/PhysRevD.74.063006.

Guido d'Amico, Jérôme Gleyzes, Nickolas Kokron, Katarina Markovic, Leonardo Senatore, Pierre Zhang, Florian Beutler, and Héctor Gil-Marín. "The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure". *JCAP*, 2020(05):005–005, May 2020. doi:10.1088/1475-7516/2020/05/005.

Dark Energy Survey Collaboration and South Pole Telescope Collaboration. "Dark Energy Survey Year 1 Results: A Precise H_0 Estimate from DES Y1, BAO, and D/H Data". *MNRAS*, 480(3):3879–3888, November 2018. doi:10.1093/mnras/sty1939.

Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. "Real-Time Gravitational Wave Science with Neural Posterior Estimation". *Phys. Rev. Lett.*, 127:241103, December 2021. doi:10.1103/PhysRevLett.127.241103.

Victoria de Sainte Agathe, Christophe Balland, Hélion du Mas des Bourboux, Nicolás G. Busca, Michael Blomqvist, Julien Guy, James Rich, Andreu Font-Ribera, Matthew M. Pieri, Julian E. Bautista, et al. "Baryon acoustic oscillations at $z = 2.34$ from the correlations of Lyα absorption in eBOSS DR14". *A&A*, 629:A85, September 2019. doi:10.1051/0004-6361/201935638.

Avishai Dekel and Ofer Lahav. "Stochastic Nonlinear Galaxy Biasing". *ApJ*, 520(1):24, July 1999. doi:10.1086/307428.

Walter Del Pozzo, Tjonnie G. F. Li, and Chris Messenger. "Cosmological inference using only gravitational wave observations of binary neutron stars". *Phys. Rev. D*, 95:043502, February 2017. doi:10.1103/PhysRevD.95.043502.

Arnaud Delaunoy, Antoine Wehenkel, Tanja Hinderer, Samaya Nissanke, Christoph Weniger, Andrew R. Williamson, and Gilles Louppe. "Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization", October 2020. arXiv:2010.12931 [astro-ph.IM].

Timothée Delubac, Julian E. Bautista, Nicolás G. Busca, James Rich, David Kirkby, Stephen Bailey, Andreu Font-Ribera, Anže Slosar, Khee-Gan Lee, Matthew M. Pieri, et al. "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars". *A&A*, 574:A59, January 2015. doi:10.1051/0004-6361/201423969.

DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, et al. "The DESI Experiment Part I: Science,Targeting, and Survey Design", 2016. arXiv:1611.00036 [astro-ph.IM].

DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, D. M. Alexander, R. Alfarsy, C. Allende Prieto, M. Alvarez, O. Alves, et al. "The Early Data Release of the

Dark Energy Spectroscopic Instrument", 2023. URL https://zenodo.org/record/7964161. arXiv:2306.06308 [astro-ph.CO].

Eleonora Di Valentino, Olga Mena, Supriya Pan, Luca Visinelli, Weiqiang Yang, Alessandro Melchiorri, David F. Mota, Adam G. Riess, and Joseph Silk. "In the Realm of the Hubble tension – a Review of Solutions". *Class. Quantum Grav.*, 38(15):153001, July 2021. doi:10.1088/1361-6382/ac086d.

Scott Dodelson and Fabian Schmidt. *Modern Cosmology*. 2020.

Scott Dodelson and Michael D. Schneider. "The effect of covariance estimator error on cosmological parameter constraints". *Phys. Rev. D*, 88:063537, September 2013. doi:10.1103/PhysRevD.88.063537.

Hélion du Mas des Bourboux, Jean-Marc Le Goff, Michael Blomqvist, Nicolás G. Busca, Julien Guy, James Rich, Christophe Yèche, Julian E. Bautista, Étienne Burtin, Kyle S. Dawson, et al. "Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at z = 2.4". *A&A*, 608:A130, December 2017. doi:10.1051/0004-6361/201731731.

Hélion du Mas des Bourboux, James Rich, Andreu Font-Ribera, Victoria de Sainte Agathe, James Farr, Thomas Etourneau, Jean-Marc Le Goff, Andrei Cuceu, Christophe Balland, Julian E. Bautista, et al. "The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests". *ApJ*, 901(2):153, October 2020. doi:10.3847/1538-4357/abb085.

Albert Einstein. "Die Feldgleichungen der Gravitation". *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 844–847, January 1915.

Daniel J. Eisenstein, Idit Zehavi, David W. Hogg, Roman Scoccimarro, Michael R. Blanton, Robert C. Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". *ApJ*, 633(2):560–574, November 2005. doi:10.1086/466512.

Daniel J. Eisenstein, Hee-Jong Seo, and Martin White. "On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter". *ApJ*, 664(2):660, August 2007. doi:10.1086/518755.

Yanan Fan, David J. Nott, and Scott A. Sisson. "Approximate Bayesian computation via regression density estimation". *Stat*, 2(1):34–48, 2013. doi:https://doi.org/10.1002/sta4.15.

James Farr, Andreu Font-Ribera, Hélion du Mas des Bourboux, Andrea Muñoz-Gutiérrez, F. Javier Sánchez, Andrew Pontzen, Alma Xochitl González-Morales, David Alonso, David Brooks, Peter Doel, et al. "LyaCoLoRe: synthetic datasets for current and future Lyman-α forest BAO surveys". *JCAP*, 2020(03):068–068, March 2020. doi:10.1088/1475-7516/2020/03/068.

Will M. Farr. "Accuracy Requirements for Empirically Measured Selection Functions". *Res. Notes AAS*, 3(5):66, May 2019. doi:10.3847/2515-5172/ab1d5f.

Nicholas Farrow, Xing-Jiang Zhu, and Eric Thrane. "The Mass Distribution of Galactic Double Neutron Stars". *ApJ*, 876(1):18, May 2019. doi:10.3847/1538-4357/ab12e3.

Stephen M. Feeney, Hiranya V. Peiris, Andrew R. Williamson, Samaya M. Nissanke, Daniel J. Mortlock, Justin Alsing, and Dan Scolnic. "Prospects for Resolving the Hubble Constant Tension with Standard Sirens". *Phys. Rev. Lett.*, 122(6), February 2019. doi:10.1103/physrevlett.122.061105.

Stephen M. Feeney, Hiranya V. Peiris, Samaya M. Nissanke, and Daniel J. Mortlock. "Prospects for Measuring the Hubble Constant with Neutron-Star–Black-Hole Mergers". *Phys. Rev. Lett.*, 126(17), April 2021. doi:10.1103/physrevlett.126.171102.

M. Fishbach, R. Gray, I. Magaña Hernandez, H. Qi, A. Sur, F. Acernese, L. Aiello, A. Allocca, M. A. Aloy, A. Amato, et al. "A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart". *ApJ*, 871(1):L13, January 2019. doi:10.3847/2041-8213/aaf96e.

Andreu Font-Ribera, Jordi Miralda-Escudé, Eduard Arnau, Bill Carithers, Khee-Gan Lee, Pasquier Noterdaeme, Isabelle Pâris, Patrick Petitjean, James Rich, Emmanuel Rollinde, Nicholas P Ross, et al. "The large-scale cross-correlation of Damped Lyman α systems with the Lyman α forest: first measurements from BOSS". *JCAP*, 2012(11):059–059, November 2012. doi:10.1088/1475-7516/2012/11/059.

Andreu Font-Ribera, David Kirkby, Nicolas Busca, Jordi Miralda-Escudé, Nicholas P. Ross, Anže Slosar, James Rich, Éric Aubourg, Stephen Bailey, Vaishali Bhardwaj, et al. "Quasar-Lyman α Forest Cross-Correlation from BOSS DR11 : Baryon Acoustic Oscillations". *JCAP*, 2014(05): 027, May 2014. doi:10.1088/1475-7516/2014/05/027.

Wendy L. Freedman, Barry F. Madore, Taylor Hoyt, In Sung Jang, Rachael Beaton, Myung Gyoon Lee, Andrew Monson, Jill Neeley, and Jeffrey Rich. "Calibration of the Tip of the Red Giant Branch". *ApJ*, 891(1):57, March 2020. doi:10.3847/1538-4357/ab7339.

Hunter Gabbard, Chris Messenger, Ik Siong Heng, Francesco Tonolini, and Roderick Murray-Smith. "Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy". *Nature Physics*, 18(1):112–117, December 2021. doi:10.1038/s41567-021-01425-7.

Samuel Gagnon-Hartman, John Ruan, and Daryl Haggard. "Debiasing standard siren inference of the Hubble constant with marginal neural ratio estimation". *MNRAS*, 520(1):1–13, 01 2023. ISSN 0035-8711. doi:10.1093/mnras/stad069.

Shanika Galaudage, Christian Adamcewicz, Xing-Jiang Zhu, Simon Stevenson, and Eric Thrane. "Heavy Double Neutron Stars: Birth, Midlife, and Death". *ApJl*, 909(2):L19, March 2021. doi:10.3847/2041-8213/abe7f6.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition, 2013.

Daniel George and E. A. Huerta. "Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data". *Physics Letters B*, 778:64–70, March 2018. doi:10.1016/j.physletb.2017.12.053.

Francesca Gerardi, Stephen M. Feeney, and Justin Alsing. "Unbiased likelihood-free inference of the Hubble constant from light standard sirens". *Phys. Rev. D*, 104:083531, October 2021. doi:10.1103/PhysRevD.104.083531.

Francesca Gerardi, Andrei Cuceu, Andreu Font-Ribera, Benjamin Joachimi, and Pablo Lemos. "Direct cosmological inference from three-dimensional correlations of the Lyman-α forest". *MNRAS*, 518(2):2567–2573, November 2022. ISSN 0035-8711. doi:10.1093/mnras/stac3257.

Francesca Gerardi, Andrei Cuceu, Benjamin Joachimi, Seshadri Nadathur, and Andreu Font-Ribera. "Optimal data compression for Lyman-α forest cosmology", 2023. arXiv:2309.13164 [astro-ph.CO].

Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. "MADE: Masked Autoencoder for Distribution Estimation". In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 881–889, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/germain15.html.

Jacob Golomb and Colm Talbot. "Hierarchical Inference of Binary Neutron Star Mass Distribution and Equation of State with Gravitational Waves". *ApJ*, 926(1):79, February 2022. doi:10.3847/1538-4357/ac43bc.

Satya Gontcho A Gontcho, Jordi Miralda-Escudé , and Nicolás G. Busca. "On the effect of the ionizing background on the Lyα forest autocorrelation function". *MNRAS*, 442(1):187–195, June 2014. doi:10.1093/mnras/stu860.

Calum Gordon, Andrei Cuceu, Jonás Chaves-Montero, Andreu Font-Ribera, Alma Xochitl González-Morales, J. Aguilar, S. Ahlen, E. Armengaud, S. Bailey, et al. "3D Correlations in the Lyman-α Forest from Early DESI Data", 2023. arXiv:2308.10950 [astro-ph.CO].

K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". *ApJ*, 622(2):759–771, April 2005. doi:10.1086/427976.

Philip Graff, Michael P. Hobson, and Anthony Lasenby. "An investigation into the Multiple Optimised Parameter Estimation and Data compression algorithm". *MNRAS: Letters*, 413(1): L66–L70, May 2011. ISSN 1745-3925. doi:10.1111/j.1745-3933.2011.01034.x.

Rachel Gray, Ignacio Magaña Hernandez, Hong Qi, Ankan Sur, Patrick R. Brady, Hsin-Yu Chen, Will M. Farr, Maya Fishbach, Jonathan R. Gair, Archisman Ghosh, et al. "Cosmological inference using gravitational wave standard sirens: A mock data analysis". *Phys. Rev. D*, 101(12), June 2020. doi:10.1103/physrevd.101.122001.

Stephen R Green and Jonathan Gair. "Complete parameter inference for GW150914 using deep learning". *Machine Learning: Science and Technology*, 2(3):03LT01, June 2021. doi:10.1088/2632-2153/abfaed.

Stephen R. Green, Christine Simpson, and Jonathan Gair. "Gravitational-wave parameter estimation with autoregressive neural network flows". *Phys. Rev. D*, 102(10), November 2020. doi:10.1103/physrevd.102.104057.

James E. Gunn and Bruce A. Peterson. "On the Density of Neutral Hydrogen in Intergalactic Space.". *ApJ*, 142:1633–1636, November 1965. doi:10.1086/148444.

Alan H. Guth. "Inflationary universe: A possible solution to the horizon and flatness problems". *Phys. Rev. D*, 23:347–356, January 1981. doi:10.1103/PhysRevD.23.347.

A. J. S. Hamilton. "Uncorrelated modes of the non-linear power spectrum". *MNRAS*, 312(2): 257–284, 02 2000. ISSN 0035-8711. doi:10.1046/j.1365-8711.2000.03071.x.

W. J. Handley, M. P. Hobson, and A. N. Lasenby. "Polychord: nested sampling for cosmology". *MNRAS: Letters*, 450(1):L61–L65, April 2015a. ISSN 1745-3925. doi:10.1093/mnrasl/slv047.

W. J. Handley, M. P. Hobson, and A. N. Lasenby. "Polychord: next-generation nested sampling". *MNRAS*, 453(4):4384–4398, September 2015b. ISSN 0035-8711. doi:10.1093/mnras/stv1911.

J. Hartlap, Patrick Simon, and P. Schneider. "Why your model parameter confidences might be too optimistic: Unbiased estimation of the inverse covariance matrix". *A&A*, 464:399, 2007. doi:10.1051/0004-6361:20066170.

J. B. Hartle. *An introduction to Einstein's general relativity*. 2003.

T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference,and prediction*. 2009.

W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". *Biometrika*, 57(1):97–109, 04 1970. ISSN 0006-3444. doi:10.1093/biomet/57.1.97.

Alan Heavens, Elena Sellentin, and Andrew Jaffe. "Extreme data compression while searching for new physics". *MNRAS*, 498(3):3440–3451, 2020. doi:10.1093/mnras/staa2589.

Alan F. Heavens, Raul Jimenez, and Ofer Lahav. "Massive lossless data compression and multiple parameter estimation from galaxy spectra". *MNRAS*, 317(4):965–972, October 2000. ISSN 0035-8711. doi:10.1046/j.1365-8711.2000.03692.x.

Norbert Henze and B Zirkler. "A class of invariant consistent tests for multivariate normality". *Communications in statistics-Theory and Methods*, 19(10):3595–3617, 1990. doi:10.1080/03610929008830400.

Catherine Heymans, Tilman Tröster, Marika Asgari, Chris Blake, Hendrik Hildebrandt, Benjamin Joachimi, Konrad Kuijken, Chieh-An Lin, Ariel G. Sánchez, Jan Luca van den

Busch, et al. "KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints". *A&A*, 646:A140, February 2021. doi:10.1051/0004-6361/202039063.

Matthew D. Hoffman and Andrew Gelman. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo", 2011. arXiv:1111.4246 [stat.CO].

Daniel E. Holz and Scott A. Hughes. "Using gravitational-wave standard sirens". *ApJ*, 629:15–22, 2005. doi:10.1086/431341.

Cullan Howlett, Antony Lewis, Alex Hall, and Anthony Challinor. "CMB power spectrum parameter degeneracies in the era of precision cosmology". *JCAP*, 2012(04):027–027, April 2012. doi:10.1088/1475-7516/2012/04/027.

Wayne Hu and Martin White. "Acoustic Signatures in the Cosmic Microwave Background". *ApJ*, 471(1):30, November 1996. doi:10.1086/177951.

Edwin Hubble. "A relation between distance and radial velocity among extra-galactic nebulae". *PNAS*, 15(3):168–173, 1929. doi:10.1073/pnas.15.3.168.

Mikhail M. Ivanov, Marko Simonović , and Matias Zaldarriaga. "Cosmological parameters from the BOSS galaxy power spectrum". *JCAP*, 2020(05):042–042, May 2020. doi:10.1088/1475-7516/2020/05/042.

J. H. Jeans. "The Stability of a Spherical Nebula". *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53, January 1902. doi:10.1098/rsta.1902.0012.

D. O. Jones, A. G. Riess, D. M. Scolnic, Y. C. Pan, E. Johnson, D. A. Coulter, K. G. Dettman, M. M. Foley, R. J. Foley, M. E. Huber, et al. "Should Type Ia Supernova Distances Be Corrected for Their Local Environments?". *ApJ*, 867(2):108, November 2018. doi:10.3847/1538-4357/aae2b9.

David O. Jones, Adam G. Riess, and Daniel M. Scolnic. "Reconsidering the Effects of Local Star Formation on Type Ia Supernova Cosmology". *ApJ*, 812(1):31, October 2015. doi:10.1088/0004-637X/812/1/31.

Nick Kaiser. "Clustering in real space and in redshift space". *MNRAS*, 227(1):1–21, 07 1987. ISSN 0035-8711. doi:10.1093/mnras/227.1.1.

Alex G. Kim. "Characterizing the Sample Selection for Supernova Cosmology". *The Open Journal of Astrophysics*, 4(1), February 2021. ISSN 2565-6120. doi:10.21105/astro.2007.11100.

David Kirkby, Daniel Margala, Anže Slosar, Stephen Bailey, Nicolás G Busca, Timothée Delubac, James Rich, Julian E Bautista, Michael Blomqvist, Joel R Brownstein, et al. "Fitting methods for baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9". *JCAP*, 2013(03):024–024, March 2013. doi:10.1088/1475-7516/2013/03/024.

Francisco-Shu Kitaura, Sergio Rodríguez-Torres, Chia-Hsun Chuang, Cheng Zhao, Francisco Prada, Héctor Gil-Marín, Hong Guo, Gustavo Yepes, Anatoly Klypin, Claudia G. Scóccola, et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release". *MNRAS*, 456(4):4156–4173, 01 2016. ISSN 0035-8711. doi:10.1093/mnras/stv2826.

Philippe Landry, Reed Essick, and Katerina Chatziioannou. "Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations". *Phys. Rev. D*, 101(12), June 2020. doi:10.1103/physrevd.101.123007.

Antony Lewis. "Efficient sampling of fast and slow cosmological parameters". *Phys. Rev. D*, 87: 103529, May 2013. doi:10.1103/PhysRevD.87.103529.

Antony Lewis and Sarah Bridle. "Cosmological parameters from CMB and other data: A Monte Carlo approach". *Phys. Rev. D*, 66:103511, November 2002. doi:10.1103/PhysRevD.66.103511.

Antony Lewis, Anthony Challinor, and Anthony Lasenby. "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models". *ApJ*, 538(2):473–476, August 2000. doi:10.1086/309179.

LIGO Scientific Collaboration. LIGO Algorithm Library - LALSuite. free software (GPL), 2018. doi:10.7935/GT1W-FZ16.

Thomas J. Loredo. "Accounting for Source Uncertainties in Analyses of Astronomical Survey Data". In *AIP Conference Proceedings*. AIP, 2004. doi:10.1063/1.1835214.

LSC-Virgo collaboration. "The LSC-Virgo white paper on gravitational wave data analysis – Science goals, status and plans, priorities (20112012 edition), LIGO Report T1100322, Virgo Report VIR-0353A-11 (October 2011)". PDF: https://dcc.ligo.org/public/0062/T1100322/003/PublicWhitePaper.pdf.

Jan-Matthis Lueckmann, Giacomo Bassetto, Theofanis Karaletsos, and Jakob H. Macke. "Likelihood-free inference with emulator networks". In *Symposium on Advances in Approximate Bayesian Inference*, 2018. arXiv:1805.09294 [stat.ML].

Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013. PDF: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.

Michele Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. Oxford University Press, October 2007. ISBN 978-0-19-857074-5. doi:10.1093/acprof:oso/9780198570745.001.0001.

K. G. Malmquist. "On some relations in stellar statistics". *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, March 1922. URL https://ui.adsabs.harvard.edu/abs/1922MeLuF.100....1M.

K. G. Malmquist. "A contribution to the problem of determining the distribution in space of the stars". *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 106:1–12, February 1925. URL https://ui.adsabs.harvard.edu/abs/1925MeLuF.106....1M/abstract.

Ilya Mandel, Will M. Farr, and Jonathan R. Gair. "Extracting distribution parameters from multiple uncertain observations with selection biases". *MNRAS*, 486(1):1086–1093, 2019. doi:10.1093/mnras/stz896.

S. Mastrogiovanni, R. Duque, E. Chassande-Mottin, F. Daigne, and R. Mochkovitch. "The potential role of binary neutron star merger afterglows in multimessenger cosmology". *A&A*, 652:A1, July 2021. doi:10.1051/0004-6361/202040229.

S. Mastrogiovanni, K. Leyde, C. Karathanasis, E. Chassande-Mottin, D. A. Steer, J. Gair, A. Ghosh, R. Gray, S. Mukherjee, and S. Rinaldi. "On the importance of source population models for gravitational-wave cosmology". *Phys. Rev. D*, 104(6), September 2021. doi:10.1103/physrevd.104.062009.

Simone Mastrogiovanni, Christos Karathanasis, Jonathan Gair, Gregory Ashton, Stefano Rinaldi, Hsiang-Yu Huang, and Gergely Dálya. "Cosmology with Gravitational Waves: A Review". *Annalen der Physik*, n/a(n/a):2200180, 2022. doi:https://doi.org/10.1002/andp.202200180.

Patrick McDonald and Daniel J. Eisenstein. "Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-α forest". *Phys. Rev. D*, 76:063009, September 2007. doi:10.1103/PhysRevD.76.063009.

Matthew McQuinn. "The Evolution of the Intergalactic Medium". *Annual Review of Astronomy and Astrophysics*, 54(1):313–362, 2016. doi:10.1146/annurev-astro-082214-122355.

Matthew McQuinn and Martin White. "On estimating Lyα forest correlations between multiple sightlines". *MNRAS*, 415(3):2257–2269, May 2011. doi:10.1111/j.1365-2966.2011.18855.x.

C. Messenger and J. Read. "Measuring a Cosmological Distance-Redshift Relationship Using Only Gravitational Wave Observations of Binary Neutron Star Coalescences". *Phys. Rev. Lett.*, 108:091101, February 2012. doi:10.1103/PhysRevLett.108.091101.

P. Meszaros. "The behaviour of point masses in an expanding cosmological substratum". *A&A*, 37:225–228, 1974. URL https://ui.adsabs.harvard.edu/abs/1974A%26A....37..225M/abstract.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. "Equation of State Calculations by Fast Computing Machines". *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. ISSN 0021-9606. doi:10.1063/1.1699114.

Daniel J. Mortlock, Stephen M. Feeney, Hiranya V. Peiris, Andrew R. Williamson, and Samaya M. Nissanke. "Unbiased Hubble constant estimation from binary neutron star mergers". *Phys. Rev. D*, 100(10), November 2019. doi:10.1103/physrevd.100.103523.

Radford Neal. "MCMC Using Hamiltonian Dynamics". In *Handbook of Markov Chain Monte Carlo*, pages 113–162. 2011. doi:10.1201/b10905. arXiv:1206.1901 [stat.CO].

Radford M. Neal. "Taking Bigger Metropolis Steps by Dragging Fast Variables", 2005. arXiv:0502099 [math.ST].

Samaya Nissanke, Daniel E. Holz, Scott A. Hughes, Neal Dalal, and Jonathan L. Sievers. "Exploring short gamma-ray bursts as gravitational-wave standard sirens". *ApJ*, 725:496–514, 2010. doi:10.1088/0004-637X/725/1/496.

Samaya Nissanke, Daniel E. Holz, Neal Dalal, Scott A. Hughes, Jonathan L. Sievers, and Christopher M. Hirata. "Determining the Hubble constant from gravitational wave observations of merging compact binaries", July 2013. arXiv:1307.2638 [astro-ph.CO].

Masamune Oguri. "Measuring the distance-redshift relation with the cross-correlation of gravitational wave standard sirens and galaxies". *Phys. Rev. D*, 93:083511, April 2016. doi:10.1103/PhysRevD.93.083511.

George Papamakarios, Theo Pavlakou, and Iain Murray. "Masked Autoregressive Flow for Density Estimation", May 2017. arXiv:1705.07057 [stat.ML].

George Papamakarios, David Sterratt, and Iain Murray. "Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows". In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 837–848. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/papamakarios19a.html.

P. J. E. Peebles. *"The large-scale structure of the universe"*. 1980.

P. J. E. Peebles and J. T. Yu. "Primeval Adiabatic Perturbation in an Expanding Universe". *ApJ*, 162:815, December 1970. doi:10.1086/150713.

Will J. Percival and Martin White. "Testing cosmological structure formation using redshift-space distortions". *MNRAS*, 393(1):297–308, January 2009. ISSN 0035-8711. doi:10.1111/j.1365-2966.2008.14211.x.

Will J Percival, Oliver Friedrich, Elena Sellentin, and Alan Heavens. "Matching Bayesian and frequentist coverage probabilities when using an approximate data covariance matrix". *MNRAS*, 510(3):3207–3221, December 2021. ISSN 0035-8711. doi:10.1093/mnras/stab3540.

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, others, and The Supernova Cosmology Project. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". *ApJ*, 517(2):565–586, June 1999. doi:10.1086/307221.

Oliver H.E. Philcox, Mikhail M. Ivanov, Marko Simonović , and Matias Zaldarriaga. "Combining full-shape and BAO analyses of galaxy power spectra: a 1.6% CMB-independent constraint on H_0". *JCAP*, 2020(05):032–032, May 2020. doi:10.1088/1475-7516/2020/05/032.

Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, et al. "Planck 2015 results

- XIII. Cosmological parameters". *A&A*, 594:A13, September 2016. doi:10.1051/0004-6361/201525830.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, et al. "Planck 2018 results. VI. Cosmological parameters". *A&A*, 641:A6, September 2020. doi:10.1051/0004-6361/201833910.

Andrew Pontzen and Fabio Governato. "Cold dark matter heats up". *Nature*, 506(7487):171–178, February 2014. doi:10.1038/nature12953.

Stephen K. N. Portillo, John K. Parejko, Jorge R. Vergara, and Andrew J. Connolly. "Dimensionality Reduction of SDSS Spectra with Variational Autoencoders". *The Astronomical Journal*, 160(1):45, June 2020. doi:10.3847/1538-3881/ab9644.

Pavlos Protopapas, Raul Jimenez, and Charles Alcock. "Fast identification of transits from light-curves". *MNRAS*, 362(2):460–468, September 2005. ISSN 0035-8711. doi:10.1111/j.1365-2966.2005.09305.x.

Geert Raaijmakers, Samaya Nissanke, Francois Foucart, Mansi M. Kasliwal, Mattia Bulla, Rodrigo Fernández, Amelia Henkel, Tanja Hinderer, Kenta Hotokezaka, Kamilė Lukošiūtė, et al. "The Challenges Ahead for Multimessenger Analyses of Gravitational Waves and Kilonova: A Case Study on GW190425". *ApJ*, 922(2):269, December 2021. doi:10.3847/1538-4357/ac222d.

César Ramírez-Pérez, Javier Sanchez, David Alonso, and Andreu Font-Ribera. "CoLoRe: fast cosmological realisations over large volumes with multiple tracers". *JCAP*, 2022(05):002, May 2022. doi:10.1088/1475-7516/2022/05/002.

César Ramírez-Pérez, Ignasi Pérez-Ràfols, Andreu Font-Ribera, M. Abdul Karim, E. Armengaud, J. Bautista, S. F. Beltran, L. Cabayol-Garcia, Z. Cai, S. Chabanier, et al. "The Lyman-α forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release", 2023. arXiv:2306.06312 [astro-ph.CO].

Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". *The Astronomical Journal*, 116(3):1009–1038, September 1998. doi:10.1086/300499.

Adam G. Riess, Lucas M. Macri, Samantha L. Hoffmann, Dan Scolnic, Stefano Casertano, Alexei V. Filippenko, Brad E. Tucker, Mark J. Reid, David O. Jones, Jeffrey M. Silverman, et al. "A 2.4% DETERMINATION OF THE LOCAL VALUE OF THE HUBBLE CONSTANT". 826(1):56, July 2016. doi:10.3847/0004-637x/826/1/56.

Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, Lucas Macri, Joel C. Zinn, and Dan Scolnic. "Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM". *ApJL*, 908(1):L6, 2021. doi:10.3847/2041-8213/abdbaf.

Adam G. Riess, Wenlong Yuan, Lucas M. Macri, Dan Scolnic, Dillon Brout, Stefano Casertano, David O. Jones, Yukei Murakami, Gagandeep S. Anand, Louise Breuval, et al. "A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^{-1} Mpc^{-1} Uncertainty from the Hubble Space Telescope and the SH0ES Team". *ApJL*, 934(1):L7, July 2022. doi:10.3847/2041-8213/ac5c5b.

M. Rigault, G. Aldering, M. Kowalski, Y. Copin, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, D. Baugh, S. Bongard, et al. "CONFIRMATION OF A STAR FORMATION BIAS IN TYPE Ia SUPERNOVA DISTANCES AND ITS EFFECT ON THE MEASUREMENT OF THE HUBBLE CONSTANT". *ApJ*, 802(1):20, March 2015. doi:10.1088/0004-637x/802/1/20.

M. Rigault, V. Brinnel, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, K. Barbary, S. Bongard, K. Boone, et al. "Strong dependence of Type Ia supernova standardization on

the local specific star formation rate". *A&A*, 644:A176, December 2020. doi:10.1051/0004-6361/201730404.

K. Riles. "Gravitational waves: Sources, detectors and searches". *Progress in Particle and Nuclear Physics*, 68:1–54, 2013. ISSN 0146-6410. doi:10.1016/j.ppnp.2012.08.001.

S. Rosswog, U. Feindt, O. Korobkin, M. R. Wu, J. Sollerman, A. Goobar, and G. Martinez-Pinedo. "Detectability of compact binary merger macronovae". *Class. Quantum Grav.*, 34 (10):104001, May 2017. doi:10.1088/1361-6382/aa68a9.

Donald B. Rubin. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician". *The Annals of Statistics*, 12(4):1151–1172, 1984. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2240995.

Vera C. Rubin and Jr. Ford, W. Kent. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". *ApJ*, 159:379, February 1970. doi:10.1086/150317.

B Sathyaprakash, M Abernathy, F Acernese, P Ajith, B Allen, P Amaro-Seoane, N Andersson, S Aoudia, K Arun, P Astone, et al. "Scientific objectives of Einstein Telescope". *Class. Quantum Grav.*, 29(12):124013, June 2012. doi:10.1088/0264-9381/29/12/124013.

B. S. Sathyaprakash and Bernard F. Schutz. "Physics, Astrophysics and Cosmology with Gravitational Waves". *Living Reviews in Relativity*, 12(1), March 2009. doi:10.12942/lrr-2009-2.

P. A. G. Scheuer. "A Sensitive Test for the Presence of Atomic Hydrogen in Intergalactic Space". *Nature*, 207(5000):963, August 1965. doi:10.1038/207963a0.

Maarten Schmidt. "Large Redshifts of Five Quasi-Stellar Sources". *ApJ*, 141:1295, April 1965. doi:10.1086/148217.

B. F. Schutz. "Determining the Hubble constant from gravitational wave observations". *Nature*, 323(6086):310–311, September 1986. doi:10.1038/323310a0.

D. Scolnic, R. Kessler, D. Brout, P. S. Cowperthwaite, M. Soares-Santos, J. Annis, K. Herner, H.-Y. Chen, M. Sako, Z. Doctor, et al. "How Many Kilonovae Can Be Found in Past, Present, and Future Survey Data Sets?". *ApJ*, 852(1):L3, December 2017. doi:10.3847/2041-8213/aa9d82.

Uroš Seljak. "Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly-α forest". *JCAP*, 2012(03):004–004, March 2012. doi:10.1088/1475-7516/2012/03/004.

Elena Sellentin and Alan F. Heavens. "Parameter inference with estimated covariance matrices". *MNRAS: Letters*, 456(1):L132–L136, December 2015. ISSN 1745-3925. doi:10.1093/mnrasl/slv190.

Hee-Jong Seo and Daniel J. Eisenstein. "Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys". *ApJ*, 598(2):720–740, December 2003. doi:10.1086/379122.

Naoki Seto and Koutarou Kyutoku. "Prospects of the local Hubble parameter measurement using gravitational waves from double neutron stars". *MNRAS*, 475(3):4133–4139, 01 2018. ISSN 0035-8711. doi:10.1093/mnras/sty090.

Christian N Setzer, Rahul Biswas, Hiranya V Peiris, Stephan Rosswog, Oleg Korobkin, and Ryan T Wollaeger and. "Serendipitous discoveries of kilonovae in the LSST main survey: maximizing detections of sub-threshold gravitational wave events". *MNRAS*, 485(3):4260–4273, February 2019. doi:10.1093/mnras/stz506.

Hongyu Shen, E A Huerta, Eamonn O'Shea, Prayush Kumar, and Zhizhen Zhao. "Statistically-informed deep learning for gravitational wave parameter estimation". *Machine Learning: Science and Technology*, 3(1):015007, November 2021. doi:10.1088/2632-2153/ac3843.

Ravi K. Sheth and Giuseppe Tormen. "Large-scale bias and the peak background split". *MNRAS*, 308(1):119–126, 09 1999. ISSN 0035-8711. doi:10.1046/j.1365-8711.1999.02692.x.

Scott A Sisson, Yanan Fan, and Mark A Beaumont. "Overview of ABC". In *Handbook of approximate Bayesian computation*, pages 3–54. Chapman and Hall/CRC, 2018. arXiv:1802.09720 [stat.CO].

Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial*. OUP Oxford, 2006.

John Skilling. "Nested Sampling". *AIP Conference Proceedings*, 735(1):395–405, 2004. doi:10.1063/1.1835238.

John Skilling. "Nested sampling for general Bayesian computation". *Bayesian Analysis*, 1(4):833 – 859, 2006. doi:10.1214/06-BA127.

Anže Slosar, Andreu Font-Ribera, Matthew M Pieri, James Rich, Jean-Marc Le Goff, Éric Aubourg, Jon Brinkmann, Nicolas Busca, Bill Carithers, Romain Charlassier, et al. "The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data". *JCAP*, 2011(09):001–001, September 2011. doi:10.1088/1475-7516/2011/09/001.

Anže Slosar, Vid Iršič, David Kirkby, Stephen Bailey, Nicolás G Busca, Timothée Delubac, James Rich, Éric Aubourg, Julian E Bautista, Vaishali Bhardwaj, et al. "Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9". *JCAP*, 2013(04): 026–026, April 2013. doi:10.1088/1475-7516/2013/04/026.

Stan Development Team. "PyStan: the Python interface to Stan, Version 2.17.1.0.", 2018. URL `http://mc-stan.org`.

A.A. Starobinsky. "A new type of isotropic cosmological models without singularity". *Physics Letters B*, 91(1):99–102, 1980. ISSN 0370-2693. doi:https://doi.org/10.1016/0370-2693(80)90670-X.

Ling Sun, Evan Goetz, Jeffrey S Kissel, Joseph Betzwieser, Sudarshan Karki, Aaron Viets, Madeline Wade, Dripta Bhattacharjee, Vladimir Bossilkov, Pep B Covas, et al. "Characterization

of systematic error in Advanced LIGO calibration". *Class. Quantum Grav.*, 37(22):225008, October 2020. doi:10.1088/1361-6382/abb14e.

Colm Talbot and Eric Thrane. "Fast, flexible, and accurate evaluation of Malmquist bias with machine learning: Preparing for the pending flood of gravitational-wave detections", December 2020. arXiv:2012.01317 [gr-qc].

Andy Taylor and Benjamin Joachimi. "Estimating cosmological parameter covariance". *MNRAS*, 442(3):2728–2738, June 2014. ISSN 0035-8711. doi:10.1093/mnras/stu996.

Stephen R. Taylor, Jonathan R. Gair, and Ilya Mandel. "Cosmology using advanced gravitational-wave detectors alone". *Phys. Rev. D*, 85:023535, January 2012. doi:10.1103/PhysRevD.85.023535.

The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al. "Population Properties of Compact Objects from the Second LIGO–Virgo Gravitational-Wave Transient Catalog". *ApJL*, 913(1):L7, May 2021. doi:10.3847/2041-8213/abe949.

Vaibhav Tiwari. "Estimation of the sensitive volume for gravitational-wave source populations using weighted Monte Carlo integration". *Class. Quantum Grav.*, 35(14):145009, June 2018. doi:10.1088/1361-6382/aac89d.

Jesús Torrado and Antony Lewis. "Cobaya: Bayesian analysis in cosmology". Astrophysics Source Code Library, record ascl:1910.019, October 2019. ascl:1910.019.

Jesús Torrado and Antony Lewis. "Cobaya: code for Bayesian analysis of hierarchical physical models". *JCAP*, 2021(05):057, May 2021. doi:10.1088/1475-7516/2021/05/057.

Tilman Tröster, Ariel. G. Sánchez, Marika Asgari, Chris Blake, Martín Crocce, Catherine Heymans, Hendrik Hildebrandt, Benjamin Joachimi, Shahab Joudaki, Arun Kannawadi, Chieh-An Lin, and Angus Wright. "Cosmology from large-scale structure: Constraining ΛCDM with BOSS". *A&A*, 633:L10, January 2020. doi:10.1051/0004-6361/201936772.

Benigno Uria, Marc-Alexandre Côté, Karol Gregor, Iain Murray, and Hugo Larochelle. "Neural autoregressive distribution estimation". *The Journal of Machine Learning Research*, 17(1): 7184–7220, 2016. arXiv:1605.02226 [cs.LG].

Licia Verde, Tommaso Treu, and Adam G. Riess. "Tensions between the early and late Universe". *Nature Astronomy*, 3(10):891–895, September 2019. doi:10.1038/s41550-019-0902-0.

Doğa Veske, Imre Bartos, Zsuzsa Márka, and Szabolcs Márka. "Characterizing the Observation Bias in Gravitational-wave Detections and Finding Structured Population Properties". *ApJ*, 922(2):258, December 2021. doi:10.3847/1538-4357/ac27ac.

Matt Visser. "Jerk, snap and the cosmological equation of state". *Class. Quantum Grav.*, 21(11): 2603–2615, April 2004. doi:10.1088/0264-9381/21/11/006.

Salvatore Vitale and Hsin-Yu Chen. "Measuring the Hubble Constant with Neutron Star Black Hole Mergers". *Phys. Rev. Lett.*, 121:021303, July 2018. doi:10.1103/PhysRevLett.121.021303.

Salvatore Vitale, Davide Gerosa, Will M. Farr, and Stephen R. Taylor. "Inferring the Properties of a Population of Compact Binaries in Presence of Selection Effects". In *Handbook of Gravitational Wave Astronomy*, pages 1–60. Springer Singapore, 2021. doi:10.1007/978-981-15-4702-7_45-1.

Digvijay Wadekar, Mikhail M. Ivanov, and Roman Scoccimarro. "Cosmological constraints from BOSS with analytic covariance matrices". *Phys. Rev. D*, 102:123521, 2020. doi:10.1103/PhysRevD.102.123521.

Kenneth C Wong, Sherry H Suyu, Geoff C-F Chen, Cristian E Rusu, Martin Millon, Dominique Sluse, Vivien Bonvin, Christopher D Fassnacht, Stefan Taubenberger, Matthew W Auger, et al. "H0LiCOW – XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes". *MNRAS*, 498(1):1420–1439, September 2019. doi:10.1093/mnras/stz3094.

Samantha Youles, Julian E. Bautista, Andreu Font-Ribera, David Bacon, James Rich, David Brooks, Tamara M. Davis, Kyle Dawson, Govinda Dhungana, Peter Doel, et al. "The effect of quasar redshift errors on Lyman-α forest correlation functions", 2022. arXiv:2205.06648 [astro-ph.CO].